Virologisches Institut
der Vetsuisse-Fakultät Universität Zürich

Direktor: Prof. Dr. sc. nat. ETH Cornel Fraefel

Detection of Potentially Commensal Viruses and Associated Bacteria in Pigs by Metagenomic Analysis

Inaugural-Dissertation

zur Erlangung der Doktorwürde der
Vetsuisse-Fakultät Universität Zürich

vorgelegt von

Charlotte Isabelle Rickli

Tierärztin
von Madiswil BE

genehmigt auf Antrag von

PD Dr. med. vet. Claudia Bachofen, Referentin
Prof. Dr. sc. nat. Matthias Schweizer, Korreferent

2020

CONTENTS

ZUSAMMENFASSUNG

Wir untersuchten Kotproben von gesunden und an Durchfall erkrankten Ferkeln um kommensale oder mutualistische Viren zu finden. Im Gegensatz zu Bakterien ist nicht viel über Kommensalismus bei Viren bekannt, obwohl dieser wahrscheinlich existiert. Zuerst wurden je 50 Kotproben von gesunden und an Durchfall erkrankten, unter acht Wochen alten Ferkeln per NGS untersucht, um Viren zu finden, welche positiv mit Gesundheit und kommensalen Bakterien, und negativ mit bekannten Pathogenen korrelieren. Die Kandidaten sowie ein pathogenes Virus wurden dann per (RT-)qPCR quantifiziert.

Wir fanden 27 Virus-Genera, wobei eine höhere virale Diversität in gesunden als kranken, sowie in älteren verglichen mit jüngeren Ferkeln ersichtlich war. Fünf Viren (Kobu-, Ungulate Bocaparvo-, stool-associated circular ssDNA virus, Adeno-associated und Porprismacovirus) wurden als Kandidaten ausgewählt und per (RT-)qPCR weiter analysiert, sowie Rotavirus A als Pathogen. Kobu-, Ungulate Bocaparvovirus 5, Adeno-associated und Porprismacovirus zeigten signifikant höhere virale Mengen in gesunden verglichen mit kranken Ferkeln und waren mit kommensalen Bakterien korreliert, während Rotavirus A in höheren Mengen in kranken Tieren vorhanden war.

Zusammenfassend war es uns möglich, virale Kandidaten für Kommensalismus in Schweinen zu identifizieren. Ihre genauen Funktionen und Biologie, und ob sie zur Prävention/Therapie von Durchfall beim Schwein eingesetzt werden können, werden nachfolgende Studien zeigen.

Keywords: kommensale Viren, Schweine, NGS, Virom

ABSTRACT

We examined faecal samples of healthy and diarrhoeic piglets in order to detect commensal or mutualistic viruses. Contrary to bacteria, not much is known about commensalism in viruses, even though this is likely to exist. First, 50 faecal samples each of healthy and diarrhoeic piglets less than eight weeks old were screened by NGS to find viruses that correlated positively with health and commensal bacteria and negatively with known pathogens. The candidate viruses, as well as one known pathogenic virus, were then quantified by (RT-)qPCR.

We found 27 virus genera in total, with a higher viral diversity in healthy compared to diarrhoeic, as well as older compared to younger piglets. Five viruses (Kobu-, Ungulate Bocaparvo-, stool-associated circular ssDNA virus, Adeno-associated and Porprismacovirus) were chosen as candidates for commensalism and followed up by (RT-)qPCR, as was Rotavirus A as pathogenic virus. Kobu-, Ungulate Bocaparvovirus 5, Adeno-associated and Porprismacovirus showed significantly higher viral loads in healthy compared to diarrhoeic piglets and were correlated to commensal bacteria, whereas Rotavirus A showed higher viral loads in diarrhoeic animals.

In conclusion, we were able to determine candidate viruses for commensalism in pigs. Further studies will have to unravel their function and biology in order to possibly use them to prevent or treat diarrhoeic diseases in piglets.

Keywords: commensal viruses, pigs, NGS, virome

VIRUS GLOSSARY

AAV	Adeno-associated virus
CRESS DNA viruses	Circular Rep-encoding ssDNA viruses
PCV2	porcine Circovirus 2
PEDV	porcine Enteric Diarrhoea virus
pKoV	porcine Kobuvirus
pPSV	porcine associated Porprismacovirus
PRRSV	porcine reproductive and respiratory syndrome virus
pRVA	porcine Rotavirus A
pSCV	porcine stool-associated circular ssDNA virus
TGEV	Transmissible Gastroenteritis virus
UBoV	Ungulate Bocaparvovirus

1 INTRODUCTION

Historically, viruses have always been regarded as primarily pathogenic, as they were searched for mainly in diseased individuals and because of their nature as obligate intracellular parasites (Griffiths, 1999). Identifying viruses not associated with illnesses was difficult, since viruses do not have any common markers that could be used in, e.g., amplicon sequencing. Therefore, it was impossible to examine the entirety of viruses present in a sample. However, thanks to the development and progressive affordability of new high-throughput sequencing techniques in recent years, this has become feasible. Next Generation Sequencing (NGS) is based on massively paralleled sequencing, which allows the analysis of whole genomes much faster than first generation approaches, such as Sanger sequencing (Metzker, 2010). The virome, the collection of all viruses present in an organism, can be determined by sequencing the total nucleic acid content in samples and then assigning those sequences to viruses, so-called reference-based assembly, or by aligning sequencing reads against each other without a reference, so-called *de novo* assembly. This led to the discovery of a multitude of viruses, and to the detection of the fact that viral nucleic acids can be found also in healthy organisms. This would mean that these viruses are not harming the host. Such a relationship is a form of symbiosis, which is referred to as commensalism (Roossinck, 2011). If both partners, meaning virus and host, benefitted from the relationship, the interaction is called mutualism (Roossinck, 2011). A couple of examples for this type of relationship have already been described in viruses. In humans, infection with human T-lymphotropic virus type-I was associated with a reduced risk of developing gastric cancer (Arisawa et al., 2003), although the reasons for this are not quite clear. In another study, it was found that the latent infection with certain Herpesviruses in mice resulted in resistance against subsequent infection with *Listeria monocytogenes* and *Yersinia pestis* by stimulation of the innate immune system (Barton et al., 2007). Another group found that infection with murine Norovirus could restore the intestinal morphology and lymphocyte function in germ-free or antibiotics-treated mice (Kernbauer et al., 2014). The third type of symbiosis is parasitism, where one partner benefits and the other is harmed by the interaction (Roossinck, 2011). This relationship is the case for pathogenic viruses. However, the type of symbiosis can change for a given virus and its host, depending on factors like the immune status or presence of other viruses or bacteria. For example, it is clear that porcine Circovirus 2 (PCV2) causes a complex of diseases in swine (Segalés et al., 2005, Meng, 2013).

However, it has been found that infection with the virus alone rarely resulted in disease, but that coinfections with other viruses (e.g. Parvo-, Influenza-, Torque teno virus) or even bacteria (like *Lawsonia intracellularis*) enhanced the disease (Opriessnig and Halbur, 2012). It is not exactly clear how this interaction occurs, but since PCV2 shows a tropism for lymphoid tissue, an explanation could be that the upregulation of the immune system in case of a coinfection simultaneously promotes the replication of PCV2 (Opriessnig and Halbur, 2012). Cases like these could explain why mutualism has not been confirmed in viruses yet, and why different studies on the same virus can show different results regarding pathogenicity.

The pig is one of the most important sources of meat world-wide. The production has to meet the high demand, which often means high density production which is challenging to animal well-being and health. Diarrhoea, mostly caused by infectious agents, is one of the most important factors for losses in pig production and application of medication such as antibiotics. Bacteria, such as *Escherichia coli*, *Clostridium perfringens type C* and *Salmonella*, as well as viruses (porcine Rotavirus, Coronavirus (PEDV and TGEV), Circovirus) are the most common causes of diarrhoea in piglets (thepigsite.com, 2019). The important role of commensal bacteria in the intestine has been well researched and established, and studies have even been conducted to examine the effect of feeding probiotic bacteria to pigs (Giang et al., 2012, Sachsenröder et al., 2014). However, it is not known whether viruses can also have commensal or beneficial functions in the intestine of pigs.

With increasing antibiotic resistances, new treatment options and approaches need to be established which requires more knowledge of the components and composition of the intestinal ecosystem, including the role of enteric viruses. The best way to study the entirety of viruses in an ecosystem, the virome, is metagenomic analysis using NGS. This method enables the unspecific detection of viral nucleic acids present in a sample and can therefore provide an overview of the spectrum present in the intestine of the piglet without having to target specific viruses. So far, several studies have already looked into the composition of the porcine enteric or faecal virome. In Europe, Kobu- and CRESS DNA viruses were most often found (Sachsenröder et al., 2012, Karlsson et al., 2016). Studies from the USA showed mainly Kobu- and Sapoviruses in healthy, and Astro- and Enteroviruses in diarrhoeic animals (Shan et al., 2011, Lager et al., 2012). In China, Kobuviruses were most often found in healthy piglets, whereas Enteroviruses were the most frequent in diarrhoeic animals (Zhang et al., 2014). In healthy piglets in East Africa, Astroviruses were the most abundant (Amimo et al., 2016). These

findings show that the most frequent viruses not only vary depending on the health status and the age of the animals, but also on the country. Differences in husbandry conditions, like high-density farming in the USA and China, and the epizootic status play a role in the composition of the enteric virome. Switzerland for example is free of epizootics like Aujeszky's disease (Suid Herpesvirus 1), PEDV, TGEV or PRRSV, and the Swiss swine population is quite isolated from other countries. Also, there are hardly any high-density conditions comparable to the situation e.g. in China, and the general health status is considered to be quite good. This will have an influence on the viruses that circulate in Swiss pigs. However, no studies on the enteric or faecal virome in pigs in Switzerland are available so far.

The biology and role of most of the previously reported viruses are still quite unknown. A good example is porcine Kobuvirus (pKoV). According to the International Committee on Taxonomy of Viruses, this genus belongs to the family of Picornaviridae and currently consists of five species, Aichivirus A – F (ICTV Master Species List 2018b.v2, MSL #34). Aichivirus C infects pigs and was first detected in Hungary in 2007, notably in healthy piglets (Reuter et al., 2008). So far, it was not possible to successfully propagate the virus in cell cultures, and there are contradictory reports regarding its pathogenicity. An early report of pKoV from Brazil showed an association (but no verified causality) with diarrhoea (Barry et al., 2011). However in the USA, infection with pKoV was not associated with diarrhoea (Verma et al., 2013). A study from the Czech Republic examining healthy pigs found the virus in 87 % of the samples (Dufkova et al., 2013). In Thailand and Slovakia, diarrhoeic and healthy piglets were equally infected with the virus (Chuchaona et al., 2017, Jackova et al., 2017). A Belgian group could not associate pKoV with neither diarrhoea nor infection with Rotaviruses (Theuns et al., 2018). In Hungary, healthy pigs were more often positive for pKoV, but the difference was not statistically significant (Valkó et al., 2019).

In summary, the composition, but especially also the clinical and biological importance of many members of the enteric or faecal virome are still quite unknown. Further knowledge is necessary, especially regarding porcine diarrhoea and One Health approaches to animal and human health. Pigs can serve as a model for humans, since they are also omnivores and their digestive system is relatively similar to the human gut. A better knowledge about commensal viruses in pigs could therefore also be of importance for human health. Certain porcine enteric viruses are also discussed to have zoonotic potential, or have at least closely related pendants in humans, such as Rota- (Malasao et al., 2018) and Kobuviruses (Reuter et al., 2008). A better

understanding of the interplay of the members of the enteric ecosystem altogether is required to possibly improve prevention and treatment of enteric diseases in animals and humans.

1.1 Aims

The aim of this project was to analyse the faecal virome of healthy and diarrhoeic Swiss piglets in order to identify viruses that exist in a commensal or even mutual symbiosis with the porcine host, i.e. are not associated with diarrhoea or may even be associated with health. In addition, analysis of the faecal bacteriome will enable the identification of viruses that, on one hand, are associated with health and commensal bacteria, and on the other hand are negatively correlated with known pathogens. After defining the most promising candidate viruses for commensalism, specific testing by (RT-)qPCR will allow quantitative comparison of viruses present in diarrhoeic and healthy animals.

Furthermore, this study intends to contribute to the general knowledge on the viral and bacterial spectrum present in healthy and diarrhoeic piglets in Switzerland.

2 MATERIAL & METHODS

For the virome analysis, we used a previously established protocol that enriches for viral nucleic acids through filtration, nuclease treatment, reverse transcription and unspecific amplification. Since the filtration is supposed to also remove bacteria, the protocol was not suited for the bacteriome analysis, which is why we performed the commonly used and established 16S rRNA amplicon sequencing method for that part. Five candidate viruses as well as one pathogenic virus were then specifically analysed by (RT-)qPCR for quantification.

2.1 Farms and animals

In this study, faecal samples from 100 different piglets were analysed. They originated from 56 different farms in Central Switzerland, where the majority of Swiss pig producers are located. There was no information about the housing system of the animals available. At the time of sample collection, the animals were aged between three days to eight weeks.

The study population consisted of two groups of 50 animals, one group of healthy animals and one of piglets that suffered from diarrhoea. There is no significant difference in the age distribution of animals in the two groups (Figure 1 A and B).

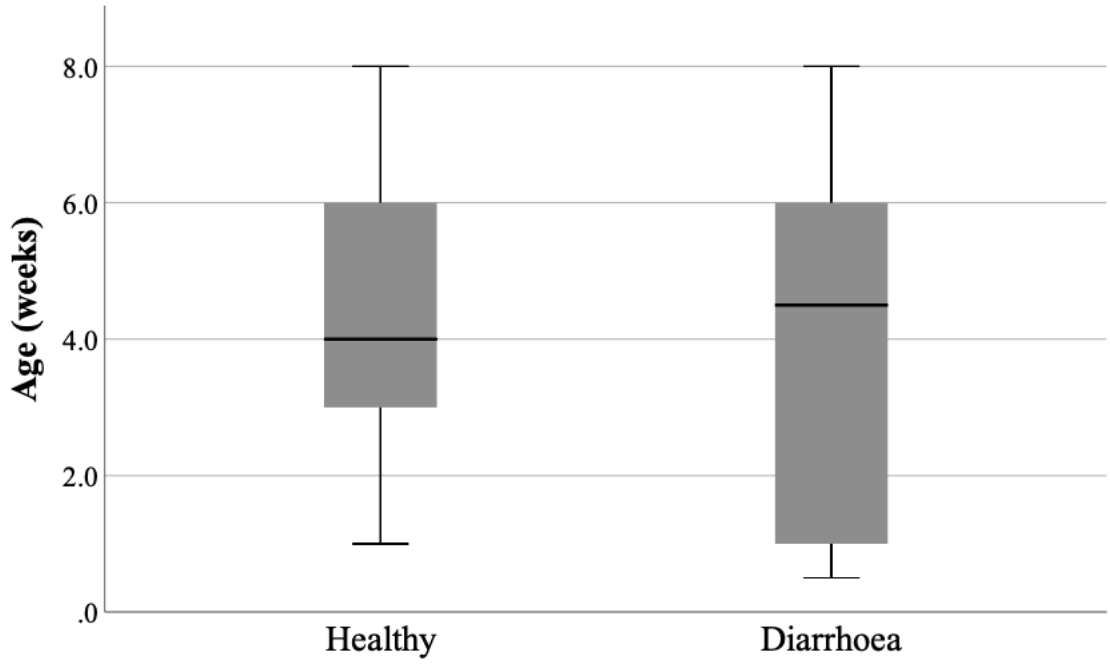

Figure 1 A: Boxplot of the distribution of the age in healthy and diarrhoeic piglets

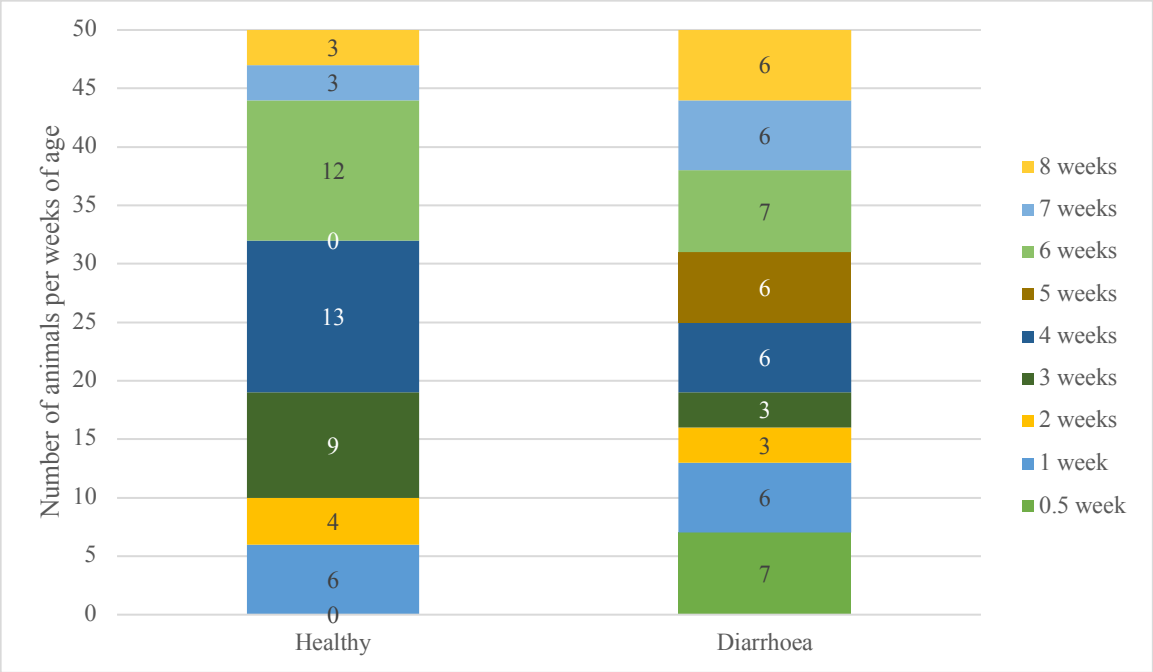

Figure 1 B: Distribution of the age in healthy and diarrhoeic piglets

In order to be included in the study, the healthy animals had to be clinically healthy without any signs of diarrhoea or other diseases. Additionally, they had to have no history of antibiotic treatment. The piglets of the other group were macroscopically diagnosed with diarrhoea during necropsy. Other signs of illness, such as septicaemia or other secondary infections were present in some cases. The cause of death was either naturally or by euthanasia.

2.2 Sample collection

2.2.1 Diarrhoeic samples

The faecal samples of diarrhoeic animals were collected between January 2017 and April 2018 from piglets that were sent to the Institute for Veterinary Pathology at the University of Zurich in the course of the PathoPig project initiated by the Federal Food Safety and Veterinary Office. The faecal samples were taken from the rectum during the necropsy.

Upon collection, the samples were stored at -20 °C until further analysis.

2.2.2 Healthy samples

Faecal samples from healthy piglets were collected directly on the farms by vets of the pig health service for routine health status monitoring during the same time frame. The age of the

healthy piglets was chosen to match the age of the diarrhoeic animals to enable comparison between the two groups. The samples were collected rectally.
Upon collection, the samples were stored at -20 °C until further analysis.

2.3 Metagenomic Sequencing

2.3.1 Nucleic acid extraction and enrichment for viral particles

For the analysis of the bacteriome, DNA was extracted using the QIAamp DNA Stool Mini Kit (Qiagen, Switzerland) according to the manufacturer's instructions.
For the virome analysis, 30 – 100 mg of faeces were weighed in a tube and the 10-fold volume of phosphate-buffered saline (PBS) as well as one stainless steel bead (5 mm, Qiagen, Switzerland) were added. The samples were then homogenized using the TissueLyser II (Qiagen, Switzerland) for 1 minute at 20 Hz. After this, the samples were centrifuged for 5 minutes at 13'000 rpm (ca 16'000 *g*). The supernatant was aspirated and pushed through a 0.45 µm syringe filter (Puradisc, 13 mm, Whatman GE Healthcare, UK). Then, a nuclease treatment was performed to eliminate exogenous nucleic acids using 134 µl of filtrate, 14 µl of Micrococcal nuclease buffer, 1 µl of Micrococcal Nuclease (New England Biolabs, UK) and 1 µl of Ribonuclease A (Sigma-Aldrich, Switzerland) in a reaction of 15 minutes at 45 °C, followed by 1 hour at 37 °C. Finally, nucleic acid extraction was performed using the QIAamp Viral RNA Mini Kit (Qiagen, Switzerland) according to the manufacturer's instructions except for omitting the addition of carrier RNA and adding 6 µl of mercaptoethanol to 560 µl of AVL buffer in order to inactivate the nucleases from the previous step. The nucleic acids (RNA and DNA) were eluted in 30 µl water and 30 µl TRIS-EDTA.

2.3.2 Bacteriome analysis

Targeted sequencing of the V3-V4 region of the bacterial 16S rRNA gene was performed using the 341f/785r primers (Klindworth et al., 2013) with attached Illumina adaptors (Table 1) in a final concentration of 0.5 µM. The PCR mixture consisted of 5 µl of sample DNA, dNTP, primers, bovine serum albumin, 5x Phusion HF buffer as well as Phusion Hot Start II High Fidelity Polymerase (Thermo Scientific, USA), topped up to 50 µl with nuclease-free water (Table 2). Cycling conditions are shown in Table 3.

Table 1: Primer sequences for the 16S V3-V4 region

Primer	Sequence
16S 341 f	5'-CTT TCC CTA CAC GAC GCT CTT CCG ATC NNN NNN NNC CTA CGG GNG GCW GCA G-3'
16S 785 r	5'-GGA GTT CAG ACG TGT GCT CTT CCG ATC NNN NNN NNG ACT ACH VGG GTA TCT AAT CC-3'

Table 2: 16S targeted amplification PCR setup

Reagent	Volume (µl)	Concentration
5x Phusion HF Buffer	10	1x
dNTP	1	0.2 mM
Primers (341f/805r)	each 2.5	0.5 µM
Sample DNA	5	variable
Bovine serum albumin	0.4	0.4 µg/µl
Phusion Hot Start II High Fidelity Polymerase	0.5	0.02 U/µl
Water	28.1	top up

Table 3: 16S targeted amplification PCR conditions

°C	Duration	Step	Cycles
98	30 sec	Activation	
98	10 sec	Denaturation	35
54	30 sec	Annealing	
72	30 sec	Extension	
72	10 min	Final extension	
4	Hold		

The products were then run on a 1.5 % agarose gel and bands of approximately 530 base pairs length were excised with scalpel blades. Gel extraction was performed using the QIAquick Gel Extraction Kit (Qiagen, Switzerland) according to the manufacturer's instructions with an eluate volume of 50 µl.

The second PCR for the introduction of the sequencing adapters and indices (NEBNext Multiplex Oligos for Illumina Set in a final concentration of 0.5 µM) was performed using the same mixture as in the first PCR with the exception of a 15 µl instead of 5 µl DNA-input and without the addition of bovine serum albumin. The cycling conditions were identical as well except for performing only 20 cycles. Again, the products were run on a 1.5 % agarose gel and bands of 630 base pairs length were excised and extracted using the QIAquick Gel Extraction Kit (Qiagen, Switzerland) with an eluate volume of 50 µl.

The molarity of the samples was then measured on an Agilent 2200 TapeStation (Agilent Technologies, USA) using the D1000 ScreenTape assay. For the two sequencing runs, 50

samples each were randomly pooled equimolarly and then denatured and diluted according to the Illumina MiSeq System Denature and Dilute Libraries Guide (Illumina, USA).
Sequencing was performed in two runs with 50 samples each on an Illumina MiSeq (Illumina, USA) with a read length of 2 x 300 base pairs (paired end) and a high output kit.

2.3.3 Virome sequencing

After the enrichment for virus particles and extraction of nucleic acids, a reverse transcription step was carried out in order to turn viral RNA into cDNA, thus including RNA viruses in the sequencing. For that purpose, the RevertAid H Minus First Strand cDNA Synthesis Kit was used with 2.5 µM of a random hexamer primer with a known tag sequence named SISPA-N (**s**equence-**i**ndependent, **s**ingle-**p**rimer **a**mplification, Table 4). For that purpose, 28.75 µl of nucleic acid was mixed with the primer and incubated for 3 minutes at 97 °C, then cooled on ice. Then, the rest of the reagents was added (Table 5) and everything was incubated according to Table 6.

Table 4: Primer sequences for SISPA-N and SISPA

Primer	Sequence
SISPA-N	5'-GTT GGA GCT CTG CAG TCA TCN NNN NN-3'
SISPA	5'-GTT GGA GCT CTG CAG TCA TC-3'

Table 5: cDNA synthesis setup

Reagent	Volume (µl)	Concentration
Sample nucleic acid	28.75	variable
Primer (SISPA-N)	1.25	2.5 µM
5x Reaction Buffer	10	1x
RiboLock RNase Inhibitor	2.5	1 U/µl
dNTP	5	1 mM
RevertAid H Minus M-MuLV Reverse Transcriptase	2.5	10 U/µl

Table 6: cDNA synthesis conditions

°C	Duration
25	10 min
42	90 min
70	5 min

After the cDNA synthesis, 1 µl of RNase H (New England Biolabs, UK) was added and incubated for 20 minutes at 37 °C to digest the RNA.

Then, a two-step Klenow reaction was performed in order to include the primer tag-sequence in all DNA-strands. Therefore, 45.5 µl of DNA from the cDNA synthesis step were mixed with SISPA-N primer, Klenow buffer and dNTP and denatured for 1 minute at 95 °C, followed by cooling on ice. Subsequently, 25 U/µl of Klenow-fragment (Thermo Scientific, USA) was added and everything was incubated for 15 minutes at 25 °C, followed by 1 hour at 37 °C. The reaction was repeated starting with the denaturation step for the second round of 2nd strand synthesis.

The products were purified using the PureLink PCR Micro Kit (Invitrogen – Thermo Scientific, USA) according to the manufacturer's instructions with an eluate volume of 12 µl.

Amplification was carried out using the complement of the tag sequence introduced by the SISPA primer in first and 2nd strand synthesis as a primer binding site for the SISPA primer (without random hexamer). Therefore, 10 µl of sample DNA were mixed with 10x PCR buffer, dNTP, SISPA primer (Table 4) and HotStarTaq DNA Polymerase (Qiagen, Switzerland), and topped up to 50 µl with nuclease-free water (Table 7). Cycling conditions are shown in Table 8.

Table 7: Amplification PCR setup

Reagent	Volume (µl)	Concentration
10x PCR Buffer	5	1x
dNTP	1	0.2 mM
Primer (SISPA)	0.4	0.8 µM
HotStarTaq DNA Polymerase	0.5	0.05 U/µl
Sample DNA	10	variable
Water	33.1	top up

Table 8: Amplification PCR conditions

°C	Duration	Step	Cycles
95	15 min	Activation	
94	30 sec	Denaturation	
58	30 sec	Annealing	18
72	1 min	Extension	
72	10 min	Final extension	

The products were then purified using the QIAquick PCR Purification Kit (Qiagen, Switzerland) according to the manufacturer's instructions with an eluate volume of 30 µl. DNA concentration was measured on the Qubit 2.0 Fluorometer (Invitrogen – Thermo Scientific, USA) using the Qubit dsDNA HS Assay Kit (Invitrogen – Thermo Scientific, USA).

Based on the Qubit measurements, 3 ng of the sample DNA were topped up to 50 µl volume with Buffer EB (10 mM Tris-Cl, pH 8.5, Qiagen, Switzerland) and then sheared on the E220 Focused-ultrasonicator (Covaris, USA) to a fragment size of 500 base pairs. Library preparation was carried out using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, UK) according to the manufacturer's instructions, without size selection and with 8 cycles in the PCR enrichment. The NEBNext Multiplex Oligos for Illumina Set was used for indexing.

The molarity was measured on the Agilent 2200 TapeStation (Agilent Technologies, USA) using the High Sensitivity D1000 ScreenTape assay. The samples were randomly pooled equimolarly and denatured and diluted according to the NextSeq System Denature and Dilute Libraries Guide (Illumina, USA).

Sequencing was performed in two runs on an Illumina NextSeq (Illumina, USA) with a read length of 2 x 150 base pairs (paired end) and a high output kit.

2.3.4 Data analysis

Bacteriome

Identification of OTUs

For the bacteriome analysis, Operational taxonomic units (OTUs) were generated with UPARSE (usearch version 10.0.024, (Edgar, 2013)) following the example and the tutorial given for paired-end Illumina data (drive5.com/uparse/). Reads were first quality-checked with FastQC (bioinformatics.babraham.ac.uk/projects/fastqc). Following removal of sequencing adapters and low-quality bases with Trimmomatic (version 0.36 with the parameters ILLUMINACLIP:adapterSeqs:2:30:10 SLIDINGWINDOW:5:15 MINLEN:100, (Bolger et al., 2014)), paired-end reads were merged and then filtered using usearch (with the parameters -fastq_maxdiffs 25 for merging and -fastq_maxee 1 for filtering, (Edgar, 2013)). Merged reads were then truncated up to the 16S primer sequences (forward: CCTACGGGNGGCWGCAG, reverse: GACTACHVGGGTATCTAATCC) and filtered for the presence of both primer sequences with a custom python script allowing up to 2 mismatches per primer. Duplicated sequences were then collapsed with usearch and the resulting unique sequences were clustered with usearch (99 % identity and with the parameter -minsize 2, (Edgar, 2013)). OTU sequences were annotated with the taxonomy data using the RDP 16S training set (version 16, http://drive5.com/sintax/rdp_16s_v16_sp.fa.gz) and the usearch -sintax command with the

parameter -sintax_cutoff set to 0.8. OTU abundances were finally obtained by counting the number of sequences (merged and filtered) matching to the OTU sequences (usearch -otutab command with default parameters, (Edgar, 2013)). OTUs annotated as chloroplast were removed to avoid a potential bias caused by plant DNA. To avoid sequencing artefacts, OTU sequences with less than 5 counts in any of the samples were removed from all further analyses.

Differential OTU abundance

Variation in OTU relative abundance was analysed with a general linear model in R with the package DESeq2 (version 1.14.1, (Love et al., 2014)) according to the factorial design with the two explanatory factors age and diarrhoea. All individual factor combinations were coded as a unique level of a combined single factor. Specific combinations of levels were then compared with linear contrasts (Neter and Wasserman, 1974). The first set of contrasts compared animals with diarrhoea to control animals within a given age (1, 2, 3, 4, 6, 7, and 8 weeks). The second set of contrasts tested for linear trends over the age of the animals. This was fitted either separate for animals with and without diarrhoea or across the average of both. Within each comparison, *P*-values were adjusted for multiple testing (Benjamini-Hochberg), and OTUs with an adjusted *P*-value (false discovery rate, FDR) below 0.01 were considered to be differentially abundant. Normalized OTU counts were calculated accordingly with DESeq2 and log2(x+1)-transformed to obtain the normalized OTU abundances. To identify OTUs that are differentially abundant depending on the presence/absent of viruses, we used a factorial design with the two explanatory factors age and diarrhoea. To account for all effects of age, AGE was fitted first as an unordered factor. Presence/absence of a specific virus (VIRUS) was fitted second. *P*-values were corrected as described above and OTUs with an FDR < 0.01 were considered to be differentially abundant.

Analysis of biodiversity indices

To characterize the overall impact of the age (AGE) and the diagnosis on the bacterial community structure, we analysed the variation in OTU richness, diversity (Shannon Index), effective richness (exponent of the Shannon index, (Magurran, 2004)), and evenness (Pielou, 1975) with a linear model. “DIA” compared animals with diarrhoea to animals without diarrhoea. To account for all effects of age, AGE was fitted as unordered factor.

Enrichment of taxa

To test for enrichment/depletion of bacterial taxa occurrences in a given set of OTUs (e.g., OTUs with significant difference in abundance between animals with and without diarrhoea),

we constructed for each taxon a contingency table with the within/outside taxon counts for the given set of OTUs and all OTUs passing the filter. We then tested for significance with Fisher's exact test. *P*-values were adjusted for multiple testing (Benjamini-Hochberg), and phyla with an adjusted *P*-value (false discovery rate, FDR) below 0.05 were considered to be significantly enriched/depleted.

Definition of a representative sample

The medoid (the "average" or representative sample of group that is equally different to all other members of the group) was extracted using the function pam() ("Partitioning Around Medoids") from the library "cluster" in R (Maechler et al., 2012) by calculating a single cluster and extracting the medoid. Spider plots were drawn in R with the package plotrix (Lemon, 2006).

Virome

Analysis of the virome data was performed using the VirDetect pipeline, which was inspired by ezVir, a pipeline that was designed to process viral sequencing data and show the results in user-friendly and customizable way (Petty et al., 2014), and was developed using R (https://www.r-project.org/). It is implemented as an App embed in ezrun (https://github.com/uzh/ezRun) and SUSHI, which manages the analysis workflow within the SGE (Sun Grid Engine) computing cluster. In detail, raw reads were per-processed using Trimmomatic (parameters: 5 prime hard trimmed: 5 bp; minTailQuality: 10; minAvgQuality: 20; minReadLength: 50) to remove adaptors and low-quality regions. Trimmed and filtered reads were mapped to the Human genome (Ensembl GRCh38.p10) to remove contamination introduced during human handling using Bowtie 2 (parameters: very-sensitive). Un-mapped reads were extracted using samtools and aligned to the host genome (Ensembl Sscrofa11.1) to remove host contamination using Bowtie 2 with the same parameter setting. Un-mapped reads were extracted again and aligned to a proprietary viral database using Bowtie 2 (parameters: -a --very-sensitive --no-mixed --no-discordant -X 1000). Mapped reads and mapped bases per viral genome are calculated using BEDTools. Viral genomes with at least 10 mapped reads were reported using R markdown (http://rmarkdown.rstudio.com/).

The samples were then checked manually for porcine virus genera only. In order to obtain consensus sequences of certain samples, assembly was performed in SeqMan NGen, Lasergene (DNASTAR, USA) with different manually curated databases, specific for the viruses of

interest. Statistical analysis was performed using IBM SPSS Statistics (version 25.0). To find correlations with or against diarrhoea, a chi-squared test was performed for every virus genus, comparing healthy to diseased animals. P-values < 0.05 were considered significant. To test the diversities, unpaired t-tests were performed.

2.4 (RT-)qPCR

2.4.1 Nucleic acid extraction

For the analysis of RNA viruses, nucleic acid was re-extracted, this time without the enrichment that was performed for the virome analysis. For this purpose, 30 – 100 mg of faeces were weighed in a tube and the 10-fold volume of PBS and a stainless-steel bead (5 mm, Qiagen, Switzerland) were added. The samples were then homogenized using the TissueLyser II (Qiagen, Switzerland) for 1 minute at 20 Hz. After this, the samples were centrifuged for 5 minutes at 13'000 rpm (ca 16'000 *g*). Nucleic acid extraction was then performed without any enrichment using 140 µl of the supernatant in the QIAamp Viral RNA Mini Kit (Qiagen, Switzerland) according to the manufacturer's instructions with an eluate volume of 60 µl.
For DNA viruses, no re-extraction was necessary since the DNA isolated for 16S sequencing had not been enriched and could therefore directly be used in the specific qPCRs.

2.4.2 Viruses for (RT-)qPCR

The six viruses that were followed up by (RT-)qPCR were chosen based on their correlation with health or diarrhoea and the number of positive animals: the three viruses with the lowest p-values (pKoV, pSCV, pPSV), one virus with a high number of positive animals (UBoV), one virus that is of special interest at our institute (AAV) and one virus that was correlated with diarrhoea (pRVA) (Table 23).

2.4.3 RT-qPCR for porcine Kobuvirus (pKoV)

For the detection of pKoV, a previously established in-house RT-qPCR protocol was performed. The sequences of the primers used are shown in Table 9. The reaction mixture consisted of 2 µl of sample RNA, TaqMan RNA-to-C_T 1-step Kit (Applied Biosystems, USA), primers and probe, and topped up to 20 µl with nuclease-free water (Table 10). The cycling conditions are shown in Table 11. All RT-qPCR runs were performed on a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems, USA). The threshold for pKoV was set to 0.3.

Table 9: Primer sequences for pKoV. Reference strain: NC_011829.1

Primer	Sequence	Position (5' UTR start)
Forward	5'-GTT GCG TGG CTG GGA ATC-3'	486
Reverse	5'-CCA GCC GCG ACT CTA TCA AG-3'	505
Probe	5'-FAM-ACG CTT GAC CAT GTA CT-MGB-3'	542

Table 10: pKoV RT-qPCR setup

Reagent	Volume (µl)	Concentration
TaqMan RT-PCR Mix (2x)	10	1x
TaqMan RT Enzyme Mix (40x)	0.5	1x
Primer f/r	each 1	1 µM
Probe	0.8	0.2 µM
Sample RNA	2	variable
Water	4.7	top up

Table 11: pKoV RT-qPCR conditions

°C	Duration	Step	Cycles
48	30 min	Reverse Transcription	
95	10 min	Activation	
95	15 sec	Denaturation	45
60	1 min	Extension	

2.4.4 RT-qPCR for porcine Rotavirus A (pRVA)

The RT-qPCR for pRVA was performed using the VetMAX™ Swine Enteric Panel (Applied Biosystems, USA) that detects besides of pRVA also the two porcine Coronaviruses TGEV and PEDV, in combination with the Path-ID™ Multiplex One-Step RT-PCR Kit (Applied Biosystems, USA). Prior to mixing the reagents, the sample RNA was heated for 3 minutes at 97 °C and then immediately cooled on ice in order to denature the double-stranded RNA. The reaction mixture consisted of 8 µl of denatured sample RNA, the two kits and nuclease-free water to top up to 25 µl reaction volume (Table 12). The cycling conditions are shown in Table 13. The threshold was set to 0.08 for pRVA.

Table 12: pRVA RT-qPCR setup

Reagent	Volume (µl)	Concentration
2x Multiplex RT-PCR Buffer	12.5	1x
10x Multiplex RT-PCR Enzyme Mix	2.5	1x
25x Swine Enteric Panel Reagent	1	1x
Sample RNA	8	variable
Water	1	top up

Table 13: pRVA RT-qPCR conditions

°C	Duration	Step	Cycles
48	10 min	Reverse Transcription	
95	10 min	Activation	
95	15 sec	Denaturation	40
60	45 sec	Extension	

2.4.5 qPCR for Ungulate Bocaparvovirus (UBoV)

Since the different species of UBoV are highly diverse, separate primers were chosen for the two most abundant species present in our samples, UBoV2 and UBoV5. Primer sequences for UBoV2 were used according to a previous publication (Zhou et al., 2018) (Table 14). For UBoV5, a new set of TaqMan primers and probe were designed using Primer Express v3.0.1 (Applied Biosystems, USA). Nine different strains from GenBank were aligned with contigs from our samples to find conserved regions suited as primer binding sites with strain NC_016031.1 as a reference (Table 14). The two sets of primers were multiplexed in a reaction setup with the Path-ID™ Multiplex One-Step RT-PCR Kit (Applied Biosystems, USA) and 2 µl of sample DNA to a reaction volume of 10 µl (Table 15). The conditions are shown in Table 16, and the threshold was set to 0.04 for UBoV2 and 0.4 for UBoV5.

Table 14: Primer sequences for UBoV. Reference strains: KJ755666.1 (UBoV2), NC_016031.1 (UBoV5)

Species	Primer	Sequence	Position
UBoV2	Forward	5'-GCC GAT TCT GAT TTT CTC GAA-3'	184
	Reverse	5'-TGA GTC CAA ACA CGC CCT TT-3'	262
	Probe	5'-VIC-CGA TCC ACC CGC CG-MGB-3'	228
UBoV5	Forward	5'-GAT GGT CAT TTG CAG GAT TTT G-3'	2158
	Reverse	5'-AGG TAA CCA CCA CAG CGA CAA-3'	2222
	Probe	5'-FAM-CTC TAC GCT CAA GGA C-MGB-3'	2182

Table 15: UBoV qPCR setup

Reagent	Volume (µl)	Concentration
2x Multiplex RT-PCR Buffer	5	1x
10x Multiplex RT-PCR Enzyme Mix	1	1x
Primer f/r	each 0.4	0.4 µM
Probe	each 0.24	0.12 µM
Sample DNA	2	variable

Table 16: UBoV qPCR conditions

°C	Duration	Step	Cycles
50	2 min		
95	10 min	Activation	
95	15 sec	Denaturation	40
60	45 sec	Extension	

2.4.6 qPCR for porcine stool-associated circular ssDNA virus (pSCV)

For pSCV, a new set of TaqMan primers and probe was designed using Primer Express v3.0.1 (Applied Biosystems, USA). Nine strains from GenBank were aligned with contigs from our samples to find conserved regions suited as primer binding sites with strain NC_017916.1 as a reference (Table 17). The reaction mixture consisted of 1 µl of sample DNA, TaqMan Universal Master Mix (Applied Biosystems, USA), primers and probe, and topped up to 10 µl reaction volume with nuclease-free water (Table 18). The cycling conditions are shown in Table 19. The threshold was set to 0.1.

Table 17: Primer sequences for pSCV. Reference strain: NC_017916.1

Species	Primer	Sequence	Position
pSCV	Forward	5'-GCG CTC AGT CCC TTC TTT CA-3'	1921
	Reverse	5'-TCT GCT GGA GTT GGA GAA CGT-3'	1942
	Probe	5'-FAM-ATG TGC CTG TCA TCG-MGB-3'	1980

Table 18: pSCV qPCR setup

Reagent	Volume (µl)	Concentration
TaqMan Universal PCR Master Mix (2x)	5	1x
Primer f/r	each 1	0.3 µM
Probe	1	0.25 µM
Sample DNA	1	variable
Water	1	top up

Table 19: pSCV qPCR conditions

°C	Duration	Step	Cycles
50	2 min	UNG	
95	10 min	Activation	
95	15 sec	Denaturation	45
60	1 min	Extension	

2.4.7 qPCR for Adeno-associated virus (AAV)

A new set of primers and probe was designed using Primer Express v3.0.1 (Applied Biosystems, USA). Contigs from our samples were aligned to find conserved regions suited as

primer binding sites. Primers were designed based on strain DQ335246.2, with a few alterations for matching our samples (Table 20). The reaction mixture and cycling conditions were identical to those for pSCV, except for the final primer concentrations (0.9 µM) (Table 18 and Table 19). The threshold was set to 0.05.

Table 20: Primer sequences for AAV. Reference strain: DQ335246.2 (with alterations)

Species	Primer	Sequence	Position
AAV	Forward	5'-GCA ACC TCG GAA AGG CAA T-3'	2538
	Reverse	5'-CAG GCC AAA AGG TTC GAG AA-3'	2596
	Probe	5'-FAM-TTT CAG GCC AAG AAG AG-MGB-3'	2558

2.4.8 qPCR for porcine associated Porprismacovirus (pPSV)

Contigs from our samples were aligned to find conserved regions suited as primer binding sites with strain NC_039071.1 as a reference, using Primer Express v3.0.1 (Applied Biosystems, USA) (Table 21). The reaction mixture and cycling conditions were identical to those for pSCV (Table 18 and Table 19). The threshold was set to 0.04.

Table 21: Primer sequences for pPSV. Reference strain: NC_039071.1

Species	Primer	Sequence	Position
pPSV	Forward	5'-CGA TGA TGC GGT CAA AGG T-3'	1173
	Reverse	5'-CAC AGG AAG ACC AGA AGC AAG A-3'	1193
	Probe	5'-FAM-CAT GGA ATC AGG CAA AT-MGB-3'	1233

2.4.9 Normalisation

Since there is no method established to normalise faeces of different composition and consistency, we tried a dilution model based on diluting samples to the same optical density in order to account for the higher percentage of fluid in diarrhoeic samples. First, the homogenised samples from chapter 2.4.1 (faeces 1:10 diluted in PBS) were further diluted 1:3 with nuclease-free water to enable measurement. Then, the OD-value (optical density) was measured for each sample on the Tecan Infinite F50 with the software i-control I.II at 492 nm wavelength. The lowest OD value was set as target for further dilutions. Since there was no material left for several samples, 30 representative samples were chosen (diarrhoeic and healthy, different ages and consistencies) and further diluted to the target OD-value (0.8). Dilution factors ranged between 1:10 to 1:300. From the such diluted samples with very similar optical densities, nucleic acid extractions were repeated as described in chapters 2.3.1 (DNA) and 2.4.1 (RNA). Finally, the (RT-)qPCRs were repeated for pKoV, pRVA, UBoV5, AAV and pPSV as described.

3 RESULTS

Faecal samples of 50 healthy and 50 diarrhoeic piglets were screened for viruses and bacteria and analysed via (RT-)qPCR to screen for viruses that are correlated with health.

3.1 Virome analysis

3.1.1 Read counts

The two metagenomic sequencing runs produced a total of 351 and 282 million raw reads, respectively. The number of raw reads per sample ranged from 1'088'577 to 10'796'218. After quality check and host and human read removal, 583'205 to 10'681'610 reads per sample were left for analysis.

3.1.2 Viruses found

In total, 27 virus genera were found, belonging to 14 different families. Bacteriophages and viruses associated with the environment or the feed of the pigs were not included in the analysis. An overview of the viruses as well as the number of positive samples is shown in Table 22. Five viruses were found in ≥ 50 % of the animals: Kobu- (83 %), Astro- (67 %), Ungulate Bocaparvo- (66 %), Entero- (63 %) and Sapelovirus (50 %). Ten viruses were found in < 10 % of the pigs, namely Toro- (9 %), Adeno- (8 %), Gemycircular- (7 %), Rotavirus H (5 %), Circo- (4 %), Pasi- (4 %), Picobirna- (3 %), Hepatitis E virus (1 %), mammalian Orthoreo- (1 %) and Polyomavirus (1 %). Since Switzerland is free of Suid Herpesvirus 1 as well as TGEV and PEDV, these viruses were not detected.

Table 22: Overview of the viruses found by metagenomic virome analysis in alphabetical order. ***Bold****: Viruses associated with health. Italic: Virus associated with diarrhoea. P-value:* χ^2 *test for comparison of number of healthy versus diarrhoeic animals*

Group/Family		Genus	Total	Healthy	Diarrhoea	P-value
Adenoviridae		Adenovirus	8	7	1	0.059
Anelloviridae		Torque teno virus	10	4	6	0.505
Astroviridae		Astrovirus	67	32	35	0.523
Caliciviridae		**Sapovirus**	**31**	**21**	**10**	**0.017**
CRESS DNA (circular rep-encoding ssDNA viruses)	Circoviridae	Circovirus	4	1	3	0.617
		po-circo-like virus	28	16	12	0.373
		stool-associated circular virus	**25**	**17**	**8**	**0.038**
	Genomoviridae	Gemycircularvirus	7	1	6	0.112
	Smacoviridae	**Porprismacovirus**	**47**	**35**	**12**	**< 0.001**
	unclassified ssDNA viruses	**stool-associated circular ssDNA virus**	**41**	**33**	**8**	**< 0.001**
Coronaviridae		Torovirus	9	6	3	0.487
Hepeviridae		Hepatitis E virus	1	0	1	1
Parvoviridae		**Adeno-associated virus**	**10**	**9**	**1**	**0.008**
		Ungulate Bocaparvovirus	**66**	**38**	**28**	**0.035**
		Parvovirus	10	4	6	0.505
Picobirnaviridae		Picobirnavirus	3	2	1	1
Picornaviridae		**Enterovirus**	**63**	**39**	**24**	**0.002**
		Kobuvirus	**83**	**50**	**33**	**< 0.001**
		Pasivirus	4	1	3	0.617
		Posavirus	**46**	**28**	**18**	**0.045**
		Sapelovirus	50	24	26	0.689
		Teschovirus	32	19	13	0.198
Polyomaviridae		Polyomavirus	1	1	0	1
Reoviridae		mammalian Orthoreovirus	1	0	1	1
		Rotavirus A	*39*	*13*	*26*	*0.008*
		Rotavirus C	12	3	9	0.065
		Rotavirus H	5	3	2	1

The mean of the ratio of the number of positive healthy divided by the number of positive diarrhoeic animals was 1.79 (± 2.08) (Table 23). This means that on average, there were 1.79-times more virus-positive samples in healthy compared to diarrhoeic animals. This ratio ranged from 0.17 for Gemycircularvirus to 9 for AAV.

Table 23: Overview of the viruses found by metagenomic virome analysis in decreasing order of the ratio of the number of healthy positive divided by diarrhoeic positive animals. ***Bold****: Viruses selected for (RT-)qPCR*

Genus	Total	Healthy	Diarrhoea	Ratio
Adeno-associated virus	**10**	**9**	**1**	**9**
Adenovirus	8	7	1	7
stool-associated circular ssDNA virus	**41**	**33**	**8**	**4.13**
Porprismacovirus	**47**	**35**	**12**	**2.92**
stool-associated circular virus	25	17	8	2.13
Sapovirus	31	21	10	2.10
Torovirus	9	6	3	2
Picobirnavirus	3	2	1	2
Enterovirus	63	39	24	1.63
Posavirus	46	28	18	1.56
Kobuvirus	**83**	**50**	**33**	**1.52**
Rotavirus H	5	3	2	1.50
Teschovirus	32	19	13	1.46
Ungulate Bocaparvovirus	**66**	**38**	**28**	**1.36**
po-circo-like virus	28	16	12	1.33
Sapelovirus	50	24	26	0.92
Astrovirus	67	32	35	0.91
Torque teno virus	10	4	6	0.67
Parvovirus	10	4	6	0.67
Rotavirus A	**39**	**13**	**26**	**0.50**
Circovirus	4	1	3	0.33
Pasivirus	4	1	3	0.33
Rotavirus C	12	3	9	0.33
Gemycircularvirus	7	1	6	0.17
Hepatitis E virus	1	0	1	0
mammalian Orthoreovirus	1	0	1	0
Polyomavirus	1	1	0	-
Mean				**1.79**

Since there is no official taxonomy for CRESS DNA viruses (circular Rep-encoding ssDNA viruses) and the nomenclature is used inconsistently in literature, we have developed our own nomenclature, based on phylogenetic relatedness of the most closely related full-length references to the viral sequences found in our samples (Figure 2). The cluster we called stool-associated circular ssDNA viruses were all very closely related. The two stool-associated circular viruses are also close to each other, but distinctively different to the other CRESS DNA viruses found in our samples. Lastly, the nomenclature of the group of Porprismacoviruses here is based on the current classification by the ICTV (Varsani and Krupovic, 2018).

Figure 2: Phylogenetic tree of the viruses of the CRESS DNA group with the nomenclature used in this thesis. Exhaustive pairwise alignment of full-length reference sequences most closely related to the viruses present in our samples and progressive assembly of alignments using neighbour-joining phylogeny. Clone Manager 9.0

Figure 3 shows the number of positive samples of the most common virus genera in healthy and diarrhoeic samples. Remarkably, all healthy piglets, but only 33 of the diarrhoeic ones, were positive for pKoV. On the other hand, while only 13 of the healthy piglets were positive for pRVA, 26 diarrhoeic animals shed this virus.

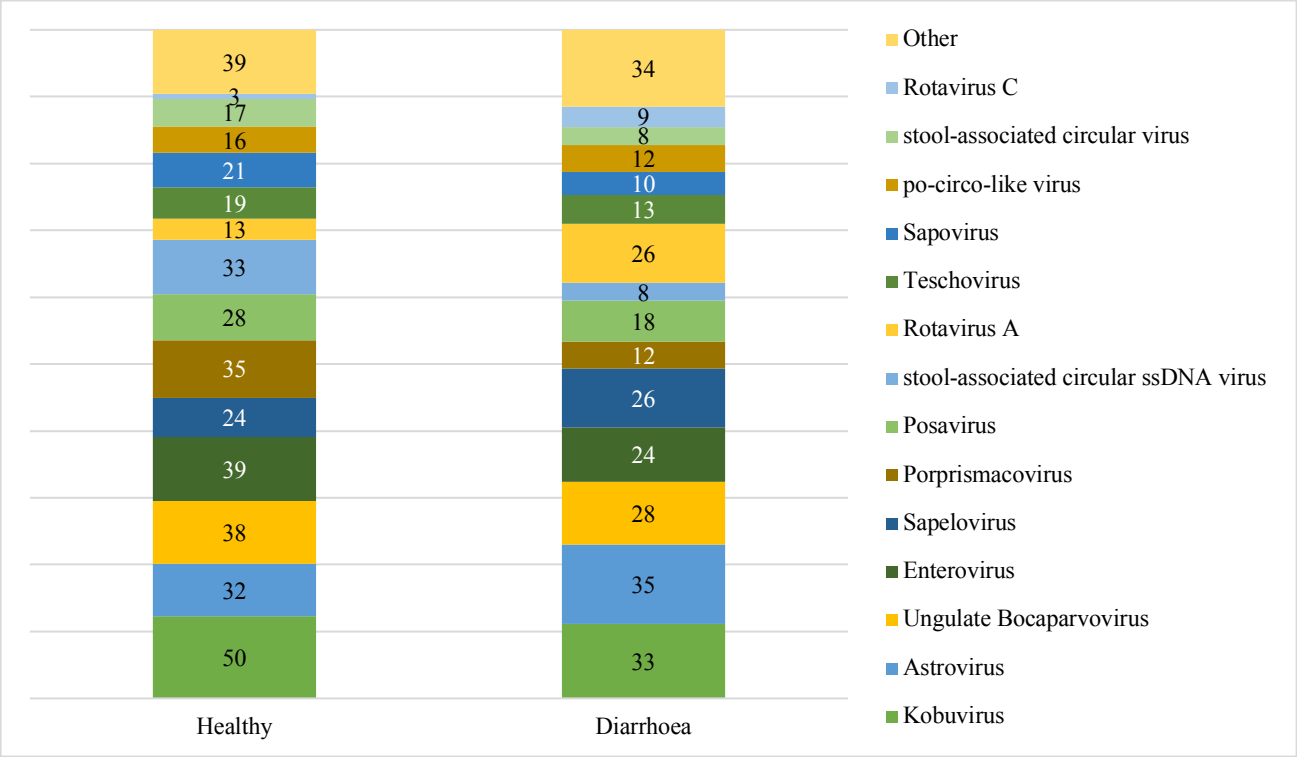

Figure 3: Number of positive samples of the 14 most frequent virus genera in healthy and diarrhoeic piglets

3.1.3 Viral diversity

The mean number of different virus genera in healthy piglets was 8.14 (± 2.7), whereas significantly less viruses were found in diarrhoeic animals with 5.92 (± 4.2) (p = 0.002) (Figure 4).

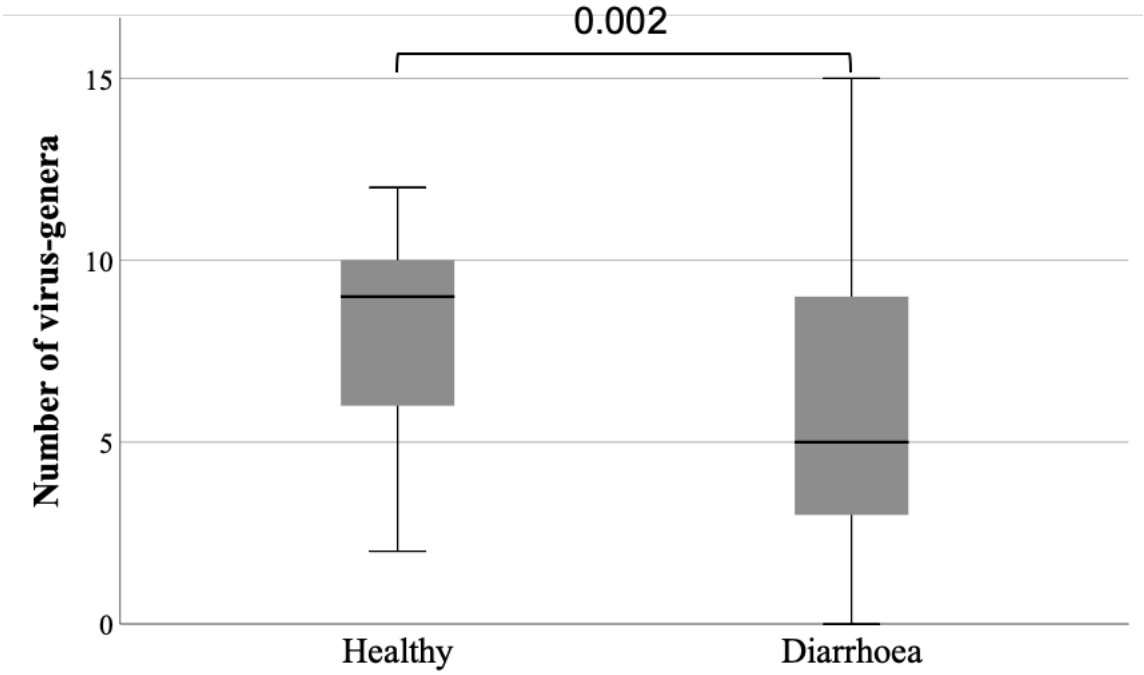

Figure 4: Boxplot of the viral diversity (number of different virus genera per sample) in healthy and diarrhoeic animals

Comparing younger (0 – 4 weeks) to older (5 – 8 weeks) animals regardless of their health status, the mean number of different genera per sample was significantly different with 5.69 (± 3.4) and 8.88 (± 3.3), respectively (p = < 0.001) (Figure 5).

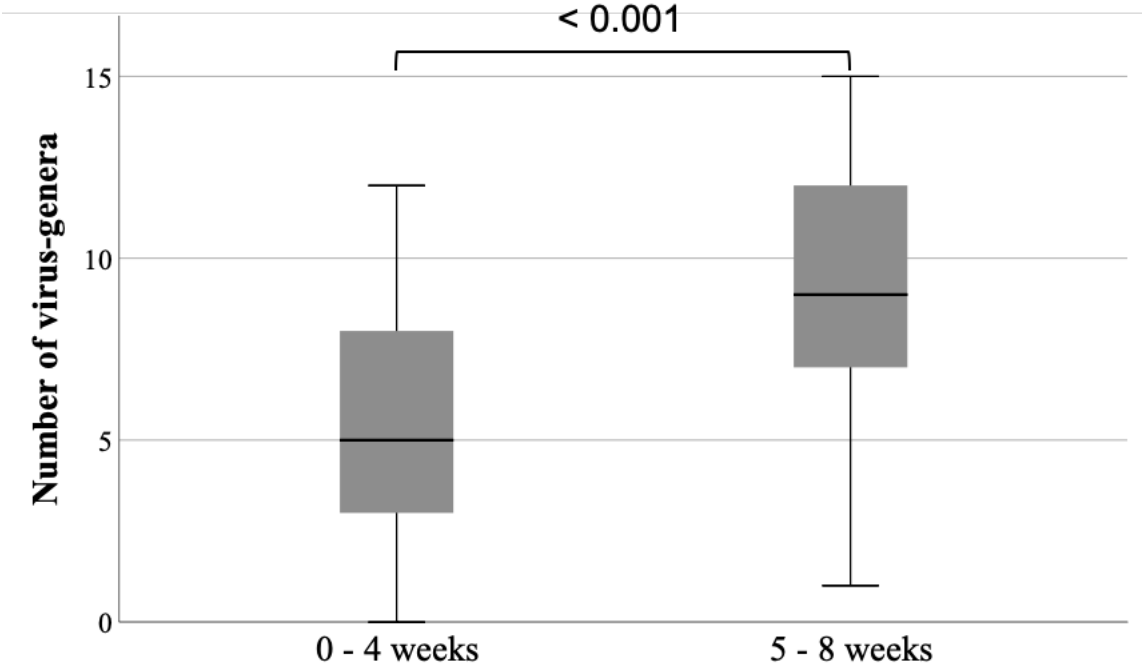

Figure 5: Boxplot of the viral diversity (number of different virus genera per sample) in younger and older animals

3.1.4 Virome of medoid samples

A medoid is a representative sample among a group that has a minimal dissimilarity to the other samples. Figure 6 shows the virome of the medoids for healthy and diarrhoeic animals. The healthy animal showed a higher diversity with nine viruses present, whereas three viruses were found in the diarrhoeic sample, with Rotavirus A being a known cause of diarrhoea. Astro- and Sapelovirus were found in both samples. Bocaparvo-, po-circo-like, Porprismaco-, Tescho-, Posa-, Kobu- and Enterovirus on the other hand were only found in the healthy piglet.

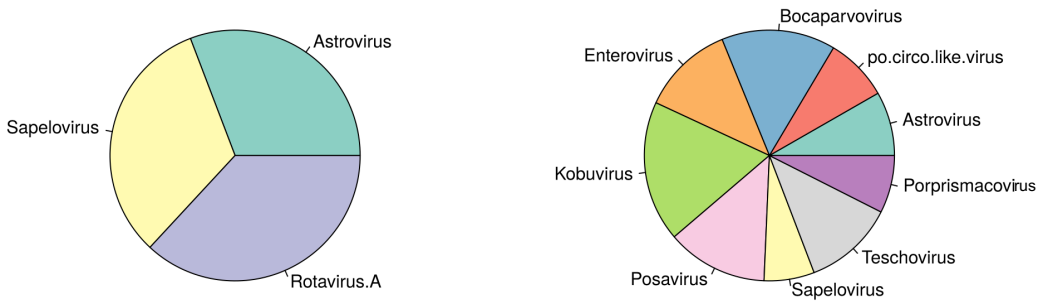

Figure 6: Medoid virome of a diarrhoeic (left) and a healthy (right) piglet

3.1.5 Candidate viruses

Nine virus genera were significantly associated with health (Sapovirus, stool-associated circular virus, Adeno-associated virus, Ungulate Bocaparvovirus, Enterovirus, Kobuvirus, Posavirus, Porprismacovirus, stool-associated circular ssDNA virus) (Table 22), whereas only pRVA was significantly associated with diarrhoea (Figure 7). Kobu-, Porprismaco- and stool-associated circular ssDNA virus had the strongest association with health, Ungulate Bocaparvovirus had a high number of positive animals, and AAV is of interest at our institute, which is why we chose to further investigate those viruses as candidates for commensalism and quantify them by q(RT-)PCR (Table 23).

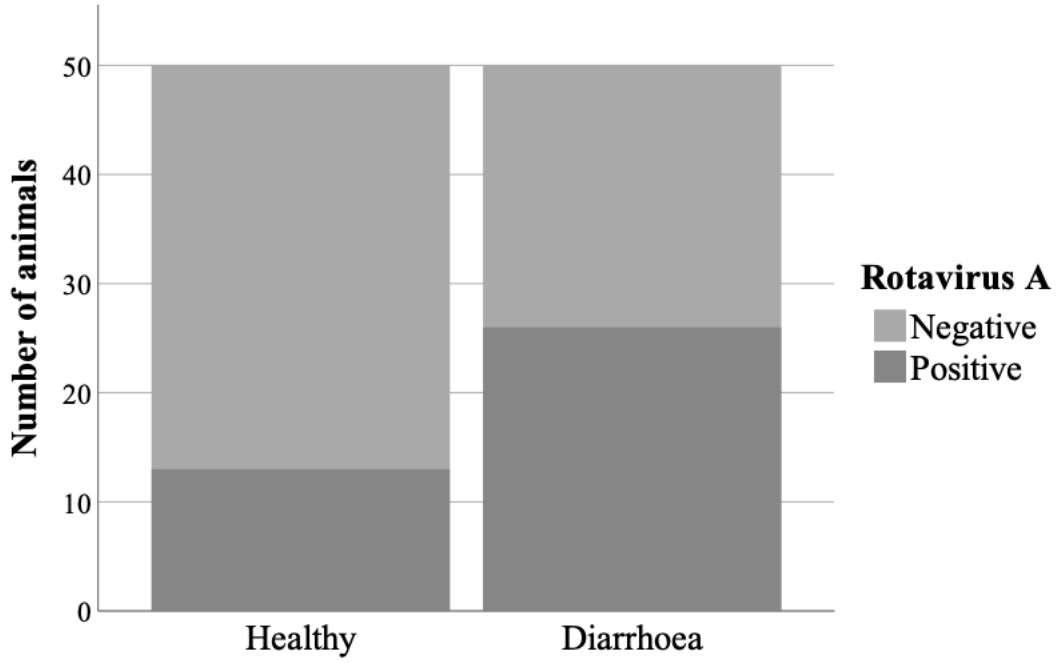

Figure 7: Number of positive and negative animals for Rotavirus A

Additionally, pPSV was significantly negatively correlated with pRVA, meaning that animals positive for pPSV were significantly more often negative for pRVA than pPSV negative animals (Table 24) ($p = 0.04$).

Table 24: Correlation of Rotavirus A with Porprismacovirus

		pPSV		
		positive	**negative**	**total**
pRVA	**positive**	13	26	39
	negative	34	27	61
	total	47	53	100

3.2 Bacteriome analysis

3.2.1 Read counts, OTUs

Both sequencing runs produced around 53 million raw reads each, with 227'000 to 650'000 raw reads per sample. After assembly and filtering, 3'575 OTUs remained for analysis.

3.2.2 Bacterial diversity indices

We tested whether diarrhoea had an influence on the biodiversity indices. The species/OTU richness (count of the number of species), Shannon index (quantitative index of species diversity), effective richness (exponent of the Shannon index) as well as Pielou's evenness (variability in species abundances) were all significantly lower in diarrhoeic animals than in the healthy ones ($p < 0.001$ for all) (Figure 8).

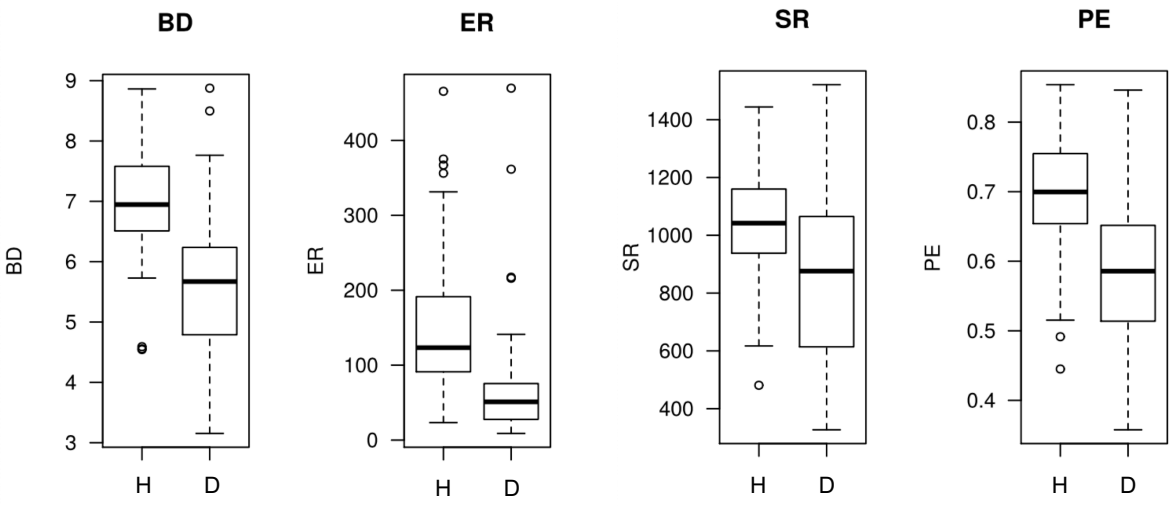

Figure 8: Diversity indices of the healthy (H) compared to the diarrhoeic (D) piglets. BD: Shannon index, ER: effective richness, SR: species richness, PE: Pielou's evenness

3.2.3 Bacteria found

In total, 3‘575 OTUs were found, belonging to 40 different families. Figure 9 gives an overview of the number of positive samples of the most common families in healthy and diarrhoeic piglets. Seven bacterial families were present in all 100 samples (Bacteroidaceae, Campylobacteraceae, Enterobacteriaceae, Fusobacteriaceae, Lactobacillaceae, Prevotellaceae, Veillonellaceae). On the other hand, Erysipelotrichaceae were present in 46 of the healthy samples and 32 of the diarrhoeic ones.

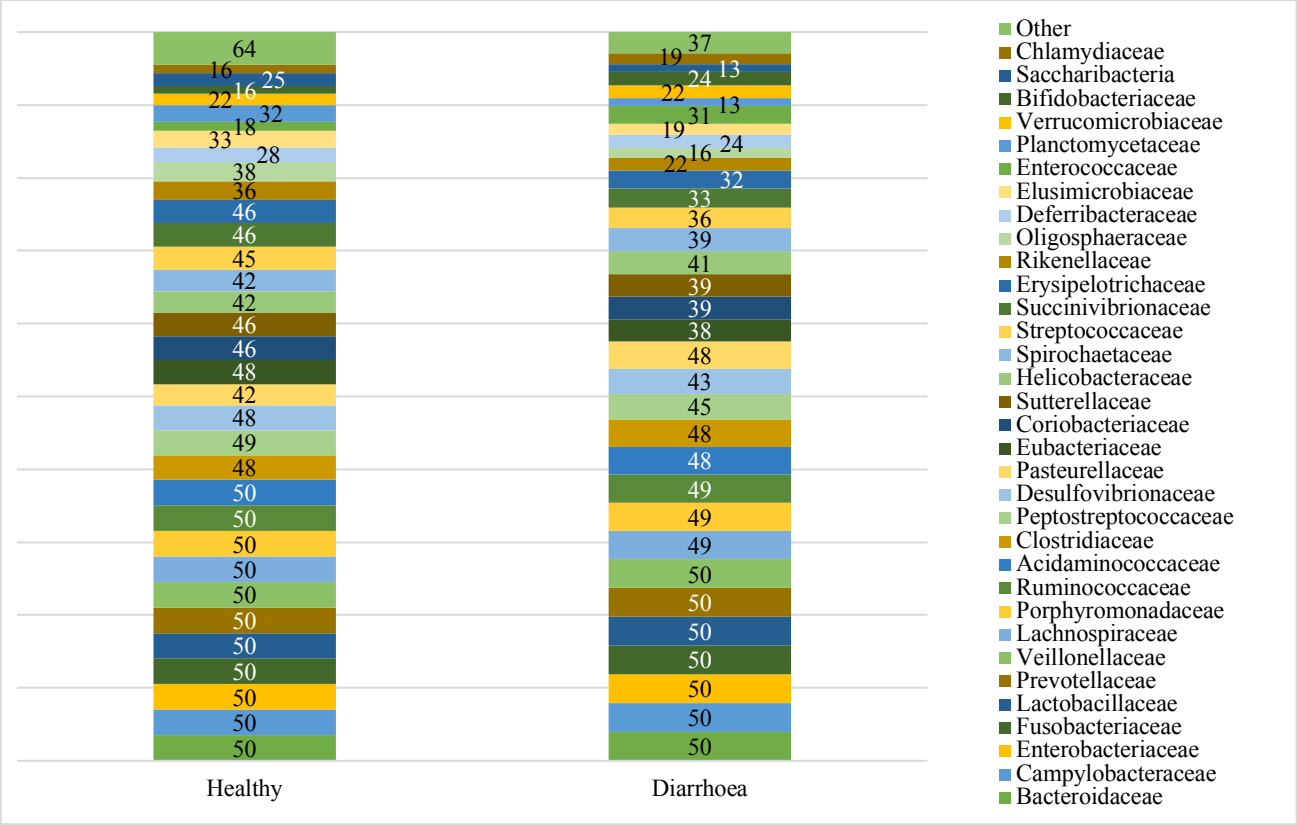

Figure 9: Number of positive samples of the most common bacterial families in healthy and diarrhoeic animals

3.2.4 Bacteriome of medoid samples

Figure 10 shows the bacteriome of the medoid samples of the healthy and the diarrhoeic group. The most abundant bacterial families in the diarrhoeic piglet were Fusobacteriaceae and Enterobacteriaceae, whereas Prevotellaceae and Ruminococcaceae were dominant in the healthy animal. Though these families could also be detected in the diarrhoeic animal, they were a lot less abundant there.

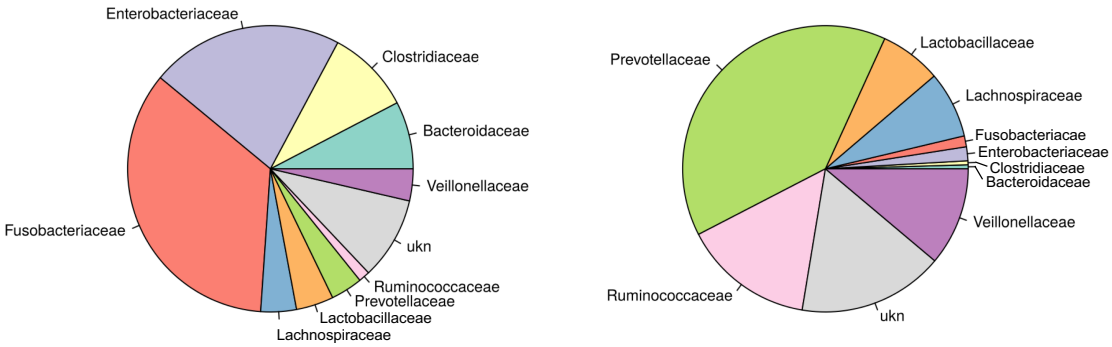

Figure 10: Medoid bacteriome of a diarrhoeic (left) and a healthy (right) piglet

3.2.5 Differential abundance of bacteria

Figure 11 shows the bacterial families that were differentially abundant in piglets with diarrhoea compared to healthy animals. In total, 483 OTUs were less abundant and 282 were more abundant in diarrhoea, resulting in a loss of OTU diversity. Most OTUs that were less abundant in diarrhoea belonged to the families of Prevotellaceae and Ruminococcaceae, whereas OTUs of Fusobacteriaceae and Enterobacteriaceae were most often more abundant in diarrhoea.

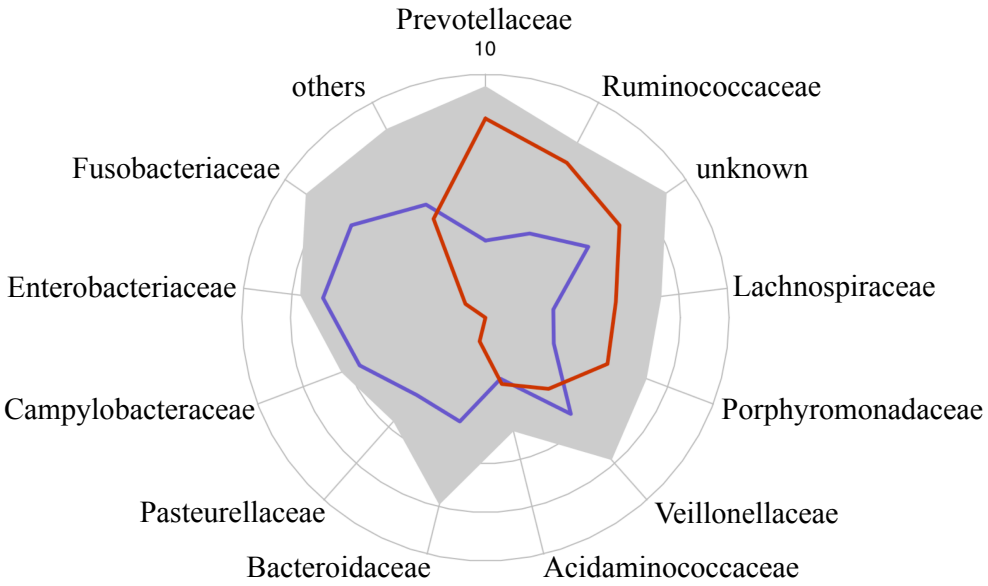

Figure 11: Spider plot of the differentially abundant OTUs in diarrhoeic compared to healthy piglets. Blue = more abundant in piglets with diarrhoea, red = less abundant in piglets with diarrhoea, grey = background distribution. log2 (x+1) with x=number OTUs

In Figure 12, the change in OTU abundance in diarrhoeic versus healthy piglets is shown split into weeks of age. Fusobacteriaceae, Bacteroidaceae, Campylobacteraceae and Enterobacteriaceae were all more abundant in diarrhoeic piglets, whereas Veillonellaceae, Prevotellaceae, Ruminococcaceae and Clostridiaceae were less abundant in diarrhoeic animals independent of the age group.

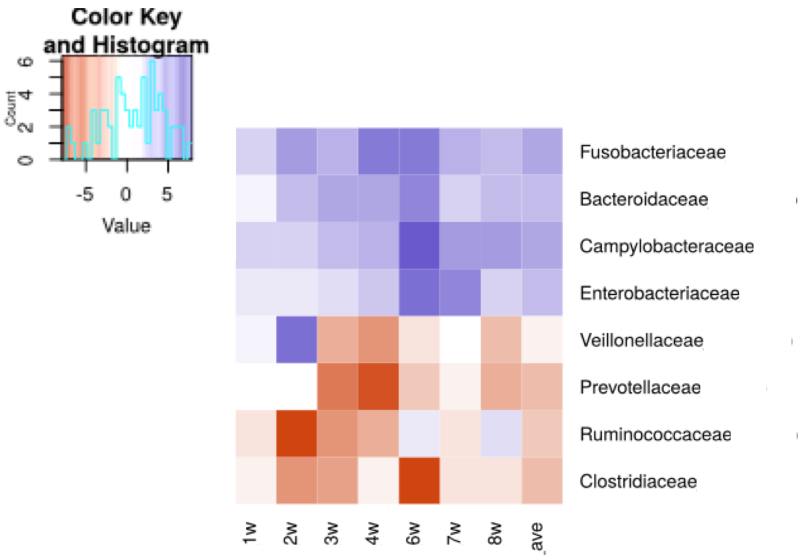

Figure 12: Heatmap of selected bacterial families and their change in abundance in diarrhoeic versus healthy piglets, split into weeks of age. The last column (ave) shows the summary of all age groups combined. Values correspond to the log2 of the number of OTUs in diarrhoeic minus the number of OTUs in healthy piglets. Red indicates a negative value, meaning a higher abundance in healthy animals, and blue indicates a positive value, meaning a higher abundance in diarrhoeic piglets

Table 25 summarises which bacterial families were associated with health over all analyses and age groups, and which ones were associated with diarrhoea.

Table 25: Overview of the bacterial families associated with health and diarrhoea

Health	**Diarrhoea**
Prevotellaceae	Fusobacteriaceae
Ruminococcaceae	Enterobacteriaceae
Lachnospiraceae	Campylobacteraceae
Porphyromonadaceae	Bacteroidaceae
Veillonellaceae	Pasteurellaceae

Figure 13 shows the difference in OTU abundances in younger compared to older piglets as a linear trend regardless of their health status. In total, 146 OTUs were less abundant, whereas 588 OTUs were more abundant in older piglets, meaning that overall, the piglets gained bacteria over the course of the first eight weeks of their life. Here, the piglets mainly gained OTUs of Prevotellaceae and Veillonellaceae as they got older, and lost OTUs of Fusobacteriaceae and Bacteroidaceae.

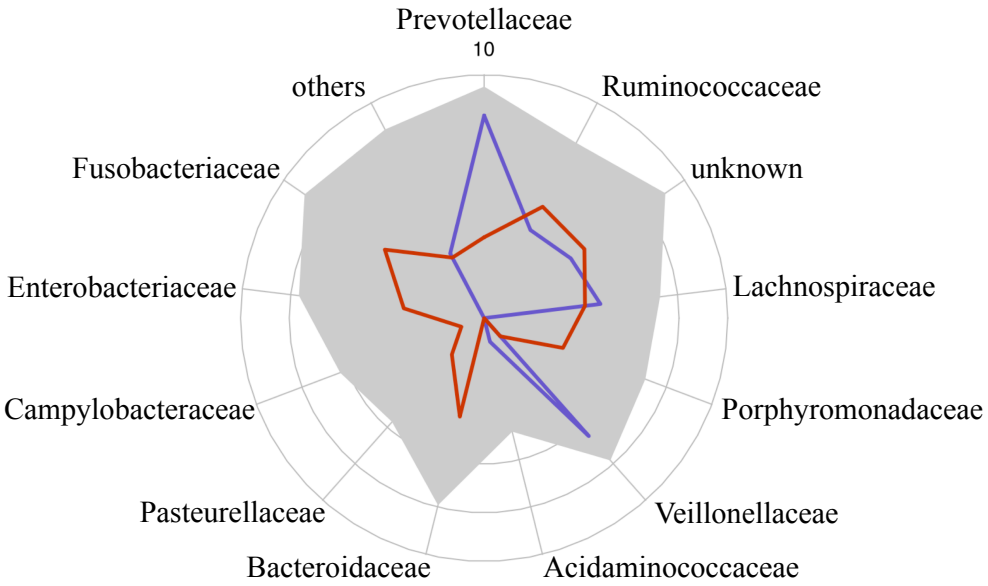

Figure 13: Spider plot of the differentially abundant OTUs in older compared to younger piglets regardless of their health status. Blue = more abundant in older piglets, red = less abundant in older piglets, grey = background distribution. log2 (x+1) with x=number of OTUs

Figure 14 shows how the families change in abundance over the weeks of age of the piglets. A clear change in the bacterial "profile" is visible around four weeks of age, which is when the piglets are weaned and change their diet. Campylobacteraceae, Veillonellaceae, Prevotellaceae and Ruminococcaceae get more abundant with progressing age, whereas Fusobacteriaceae, Bacteroidaceae, Enterobacteriaceae and Clostridiaceae are less abundant after the first couple of weeks of life.

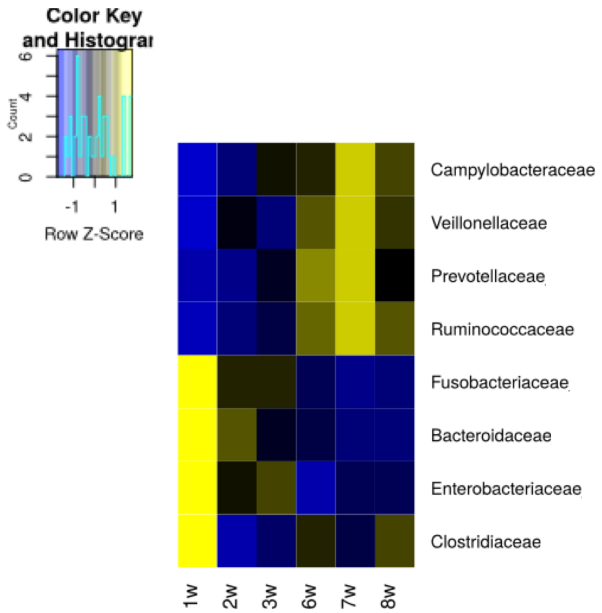

Figure 14: Heatmap of selected bacterial families and their change in abundance split in weeks of age regardless of the health condition. Values shown are virus-wise Z-scores, i.e. standardised deviations of an OTU from its average abundance across all age classes. Hence, a gradient from yellow (positive) to blue (negative) across the age classes corresponds to a decrease in abundance with age (e.g. Fusobacteriaceae). Z = xi – mean(x)/sd(x). xi: abundance of OTU x in age class i. mean(x)/sd(x): mean and standard deviation of OTU x across age classes

3.2.6 Correlation with viruses

We checked if viruses were associated with increased abundance of OTUs and if so, which bacterial OTUs were "influenced" by these viruses. Table 26 gives an overview of the number of OTUs that were more abundant if the virus was either present or absent. PSCV, UBoV and pKoV were associated with the highest numbers of differentially abundant OTUs.

Table 26: Overview of the number of OTUs that were more abundant if the respective virus was either present or absent

	Virus absent	**Virus present**	**Total**
stool-associated circular virus	361	441	802
Ungulate Bocaparvovirus	54	246	300
Kobuvirus	63	193	256
Enterovirus	96	70	166
Rotavirus C	65	88	153
Porprismacovirus	114	25	139
Pasivirus	46	21	67
Posavirus	36	23	59
po-circo-like virus	9	45	54
Rotavirus A	4	50	54
Adenovirus	25	2	27
Gemycircularvirus	14	10	24
Adeno-associated virus	7	16	23
Sapovirus	7	12	19
Circovirus	11	4	15
Torovirus	5	8	13
Rotavirus H	4	3	7
Astrovirus	1	2	3
Sapelovirus	0	3	3
Teschovirus	3	0	3
Picobirnavirus	1	0	1
Torque teno virus	0	0	0
Hepatitis E virus	0	0	0
Parvovirus	0	0	0
Polyomavirus	0	0	0
mammalian Orthoreovirus	0	0	0

In Figure 15, the number of OTUs that were associated with our candidate viruses (pSCV, UBoV, pKoV, pPSV) is pictured. Mainly Prevotellaceae and Ruminococcaceae were more abundant if the four viruses were present, whereas OTUs of Fusobacteriaceae and Enterobacteriaceae were most often more abundant in samples where the viruses were absent.

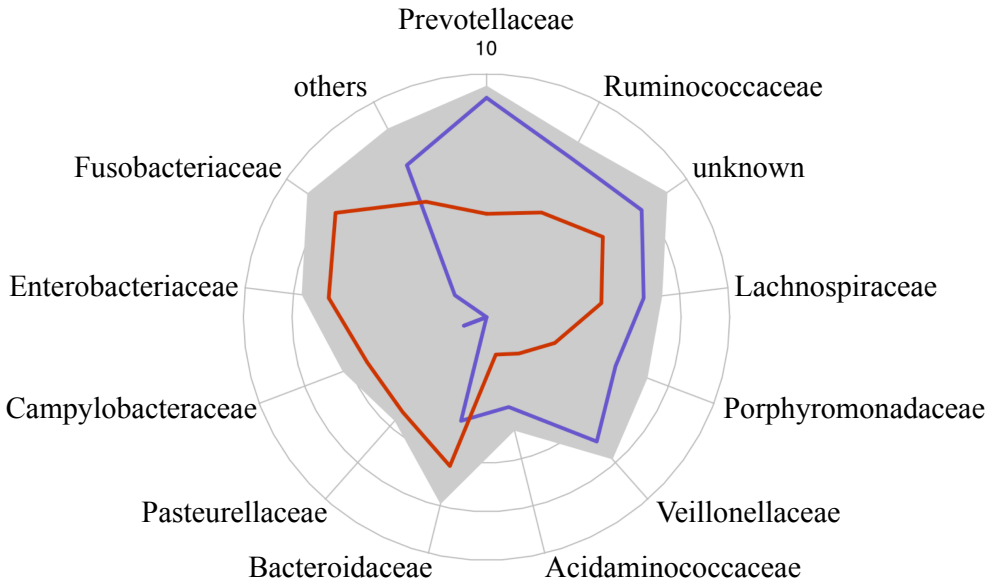

Figure 15: Spider plot of the number of OTUs that were more abundant if pSCV, UBoV, pKoV and pPSV were either present or absent. Blue = more abundant if the viruses were present, red = more abundant if the viruses were absent, grey = background distribution. log2 (x+1) with x=number of OTUs

In contrast, Figure 16 shows the number of OTUs that were more abundant if pRVA was either present or absent. We were interested to know how pRVA was associated with the abundance of bacteria since it was the only virus significantly associated diarrhoea in this study. Indeed, the presence of pRVA was mainly associated with OTUs of Campylobacteraceae, whereas OTUs of Prevotellaceae and Veillonellaceae, that were associated with health (Table 25), were more abundant if pRVA was absent.

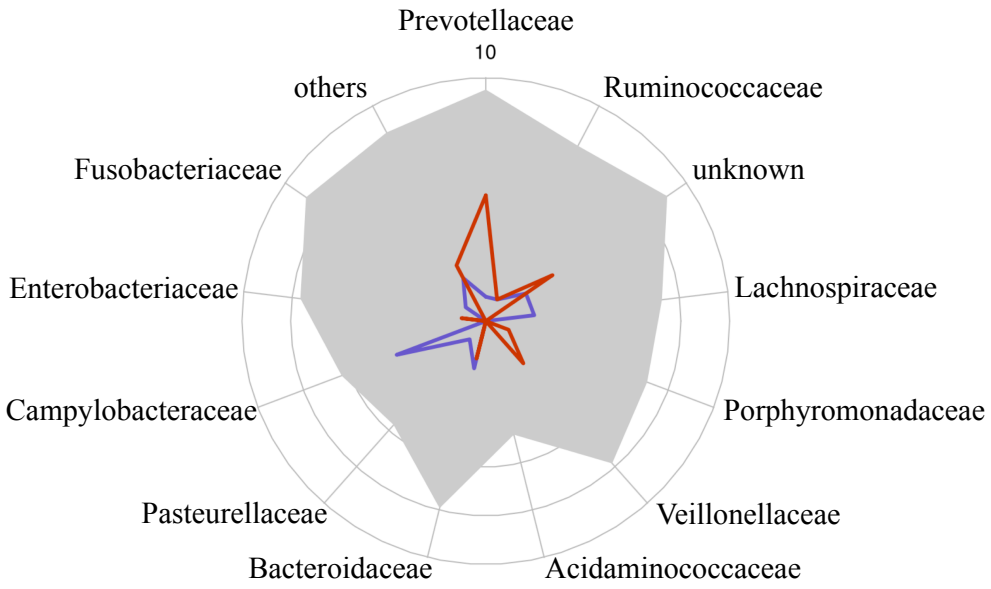

Figure 16: Spider plot of the number of OTUs that were more abundant if pRVA was either present or absent. Blue = more abundant if pRVA was present, red = more abundant if pRVA was absent, grey = background distribution. log2 (x+1) with x=number of OTUs

3.3 (RT-)qPCR

3.3.1 Porcine Kobuvirus

All samples were tested by RT-qPCR for the presence of pKoV. All healthy animals and 47 of the diarrhoeic piglets were positive, which corresponds to a prevalence of 100 % (healthy) and 94 % (diarrhoeic) in this study. The mean Ct-value for healthy animals was 22.38 (± 4.93), whereas diarrhoeic piglets showed a mean Ct-value of 26.72 (± 6.12) (Figure 17), which equals a p-value < 0.001 and therefore a significantly lower Ct-value, meaning significantly more pKoV RNA in healthy animals.

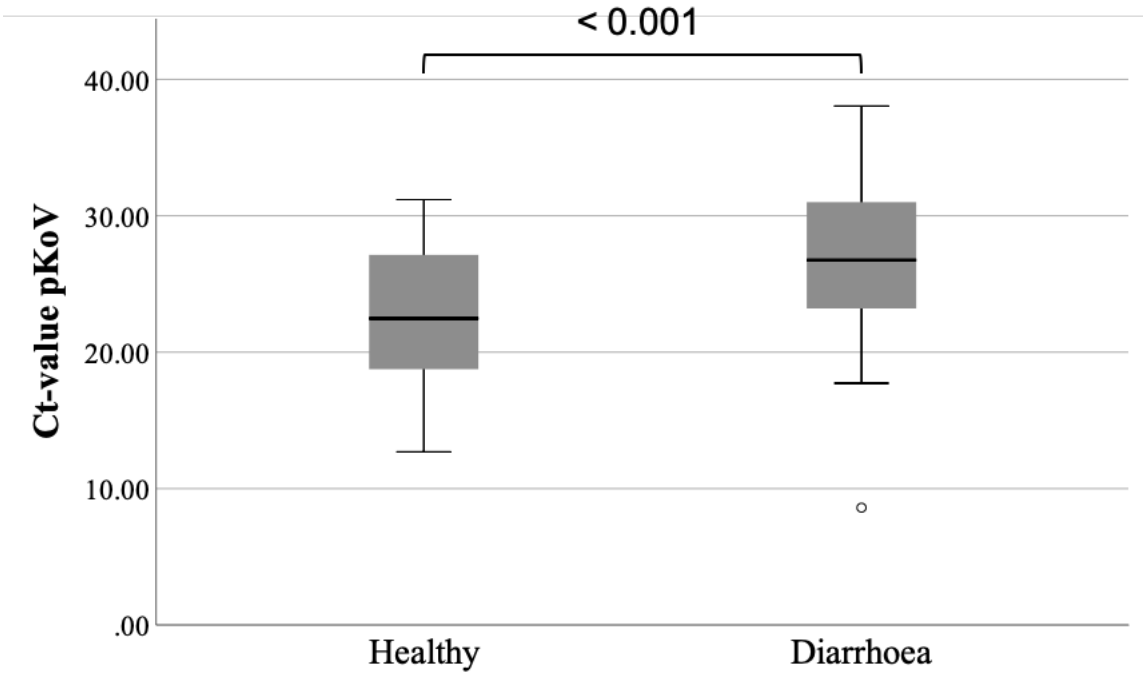

Figure 17: Boxplot of the Ct-values of pKoV in healthy and diarrhoeic piglets

3.3.2 Porcine Rotavirus A

Rotavirus A was also tested by RT-qPCR, since it was the only virus that was negatively correlated with health in sequencing. In the RT-qPCR, 35 of the healthy (70 %) and 44 (88 %) of the diarrhoeic piglets were positive for pRVA, which is a significant difference with a p-value of 0.02. The mean Ct-values were significantly ($p < 0.001$) higher in healthy (35.06 (± 5.03)) compared to diarrheic animals (27.97 (± 7.54)) (Figure 18).

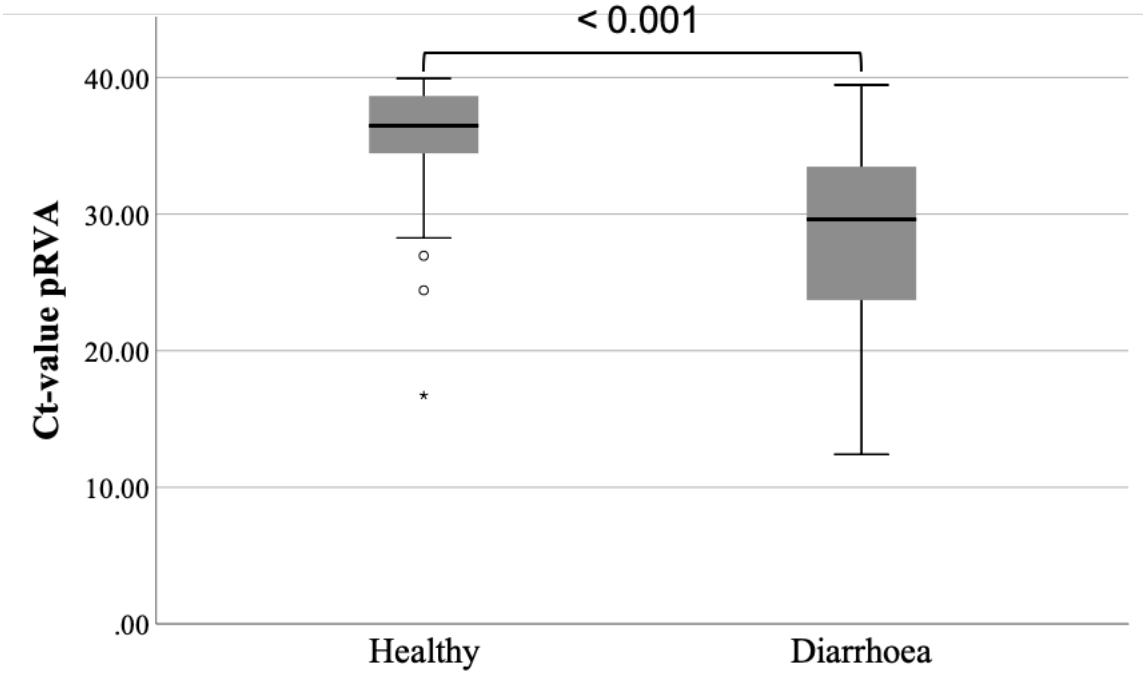

Figure 18: Boxplot of the Ct-values of pRVA in healthy and diarrhoeic piglets

3.3.3 Ungulate Bocaparvovirus

For UBoV, two sets of primers were designed to detect the two subtypes most commonly found in our samples according to the NGS data. Other subtypes were only present sporadically. For UBoV2, 43 of the healthy (86 %) and 32 of the diarrhoeic piglets (64 %) were positive, with the mean Ct-values of 28.64 (± 4.86) and 29.82 (± 4.64) being not significantly different (p = 0.291). For UBoV5, 49 (98 %) healthy and 41 (82 %) diarrhoeic animals were positive and in contrast to UBoV2, the mean Ct-value of healthy piglets was with 25.42 (± 7.2) significantly (p = 0.004) lower than in diarrhoeic animals (29.36 (± 5.34)). Figure 19 and Figure 20 show the Ct-values for the two subtypes.

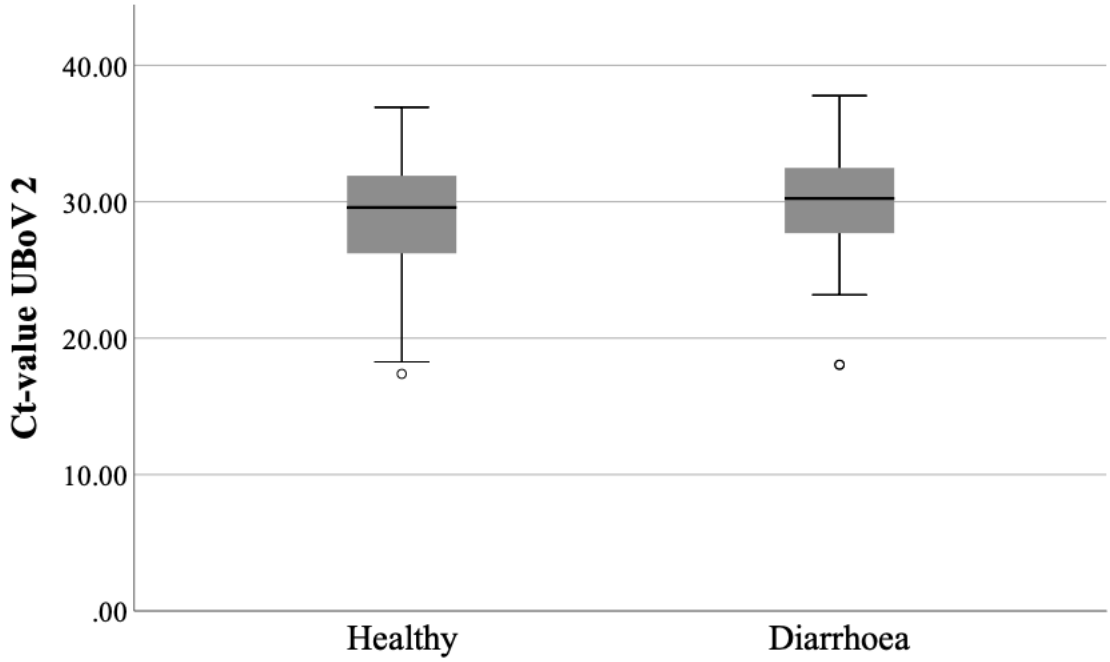

Figure 19: Boxplot of the Ct-values of UBoV2 in healthy and diarrhoeic piglets

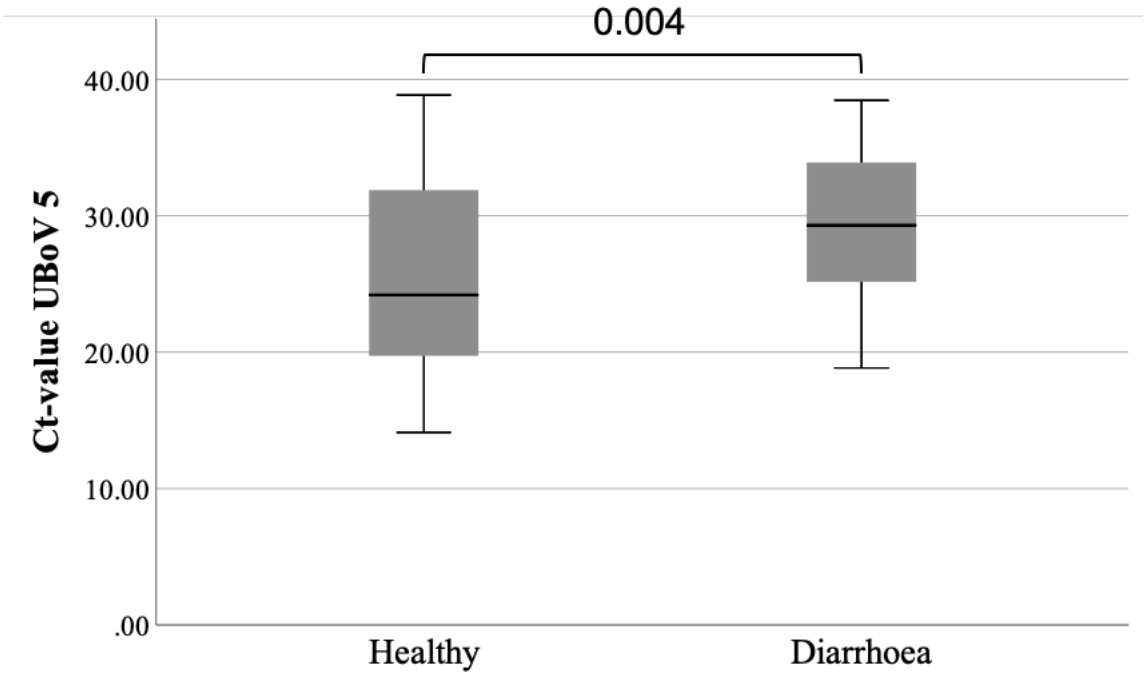

Figure 20: Boxplot of the Ct-values of UBoV5 in healthy and diarrhoeic piglets

3.3.4 Porcine stool-associated circular ssDNA virus

Of the healthy piglets, 15 (30 %) and 12 of the diarrhoeic ones (24 %) were positive for pSCV. The Ct-values were generally high and mean values were with 35.54 (± 3.43) in healthy and 36.38 (± 4.63) in diarrhoeic animals not significantly different (p = 0.594) (Figure 21).

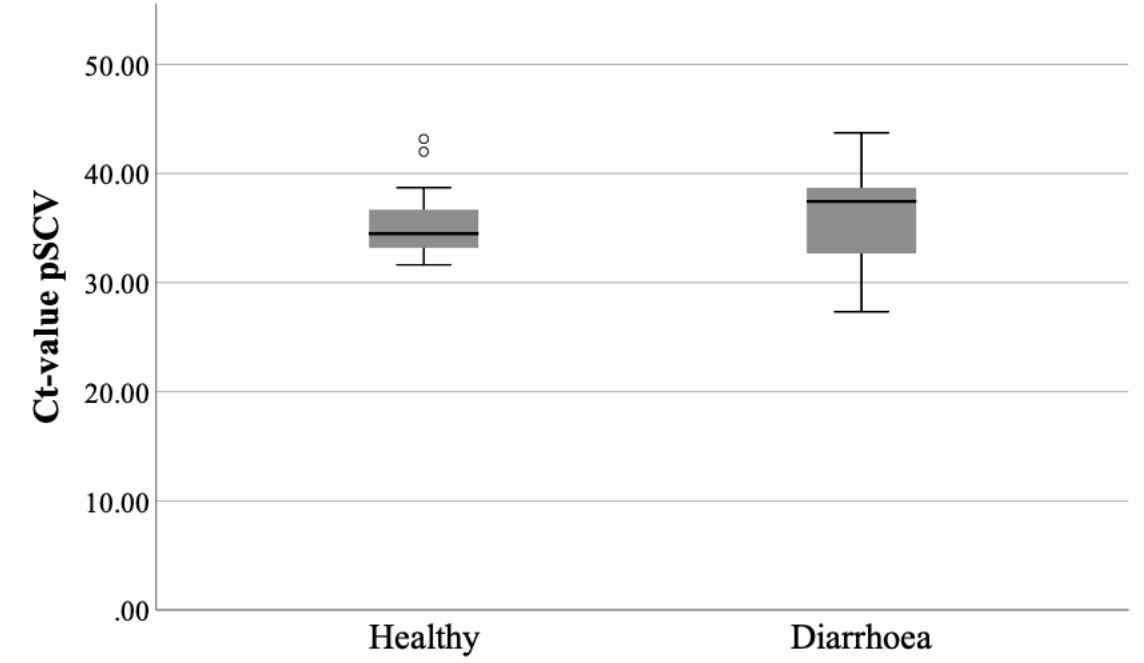

Figure 21: Boxplot of the Ct-values of pSCV in healthy and diarrhoeic piglets

3.3.5 Adeno-associated virus

While AAV was only detected in 10 animals by NGS (Table 22), qPCR revealed 35 positive samples from healthy (70 %) and 32 from diarrhoeic piglets (64 %). However, Ct-values were in many cases very high, which may explain why they were not detected by metagenomic analysis. Nevertheless, the Ct-values were significantly lower in healthy (31.13 (± 8.13)) compared to diarrhoeic animals 37.54 (± 3.19) ($p < 0.001$) (Figure 22).

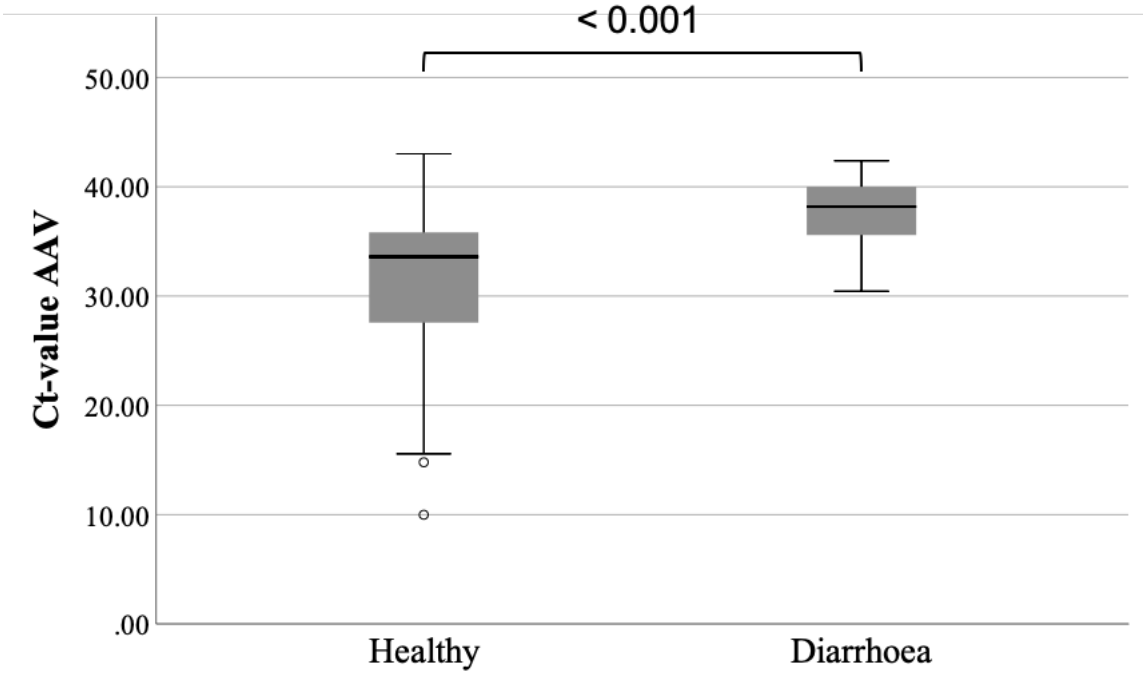

Figure 22: Boxplot of the Ct-values of AAV in healthy and diarrhoeic piglets

3.3.6 Porcine associated Porprismacovirus

Specific qPCR revealed 22 healthy (44 %) and 18 diarrhoeic (36 %) animals positive for pPSV. Also, the Ct-values were significantly (p = 0.001) lower in healthy animals (29.35 (± 5.05)) than in diarrhoeic ones (35.49 (± 5.27)) (Figure 23).

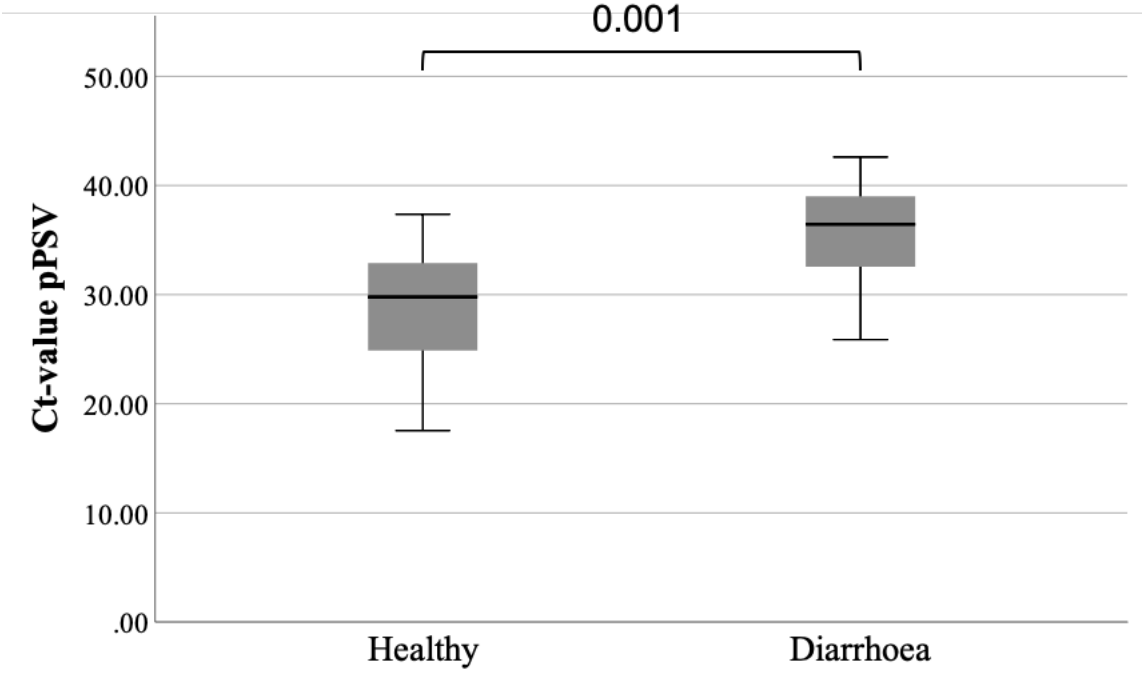

Figure 23: Boxplot of the Ct-values of pPSV in healthy and diarrhoeic piglets

3.3.7 Normalisation

For the normalisation trial, the (RT-)qPCRs with a significant difference in mean Ct-values between healthy and diarrhoeic piglets were repeated with 30 samples that were diluted to the same optical density, accounting for water influx and reduced resorption during diarrhoea. Table 27 and Figure 24 give an overview of the results. After dilution, only pRVA remained significantly different between healthy and diarrhoeic piglets. However, despite up to 300-fold dilution of samples from healthy animals, the Ct-values of the candidate viruses remained lower or equal to the values in diarrhoeic piglets. Hence, in most cases there was still more viral nucleic acid present in samples from healthy than diarrhoeic animals even if the optical density was adjusted, but the difference was not statistically significant anymore. This may at least partly be due to the fact that many samples with already high Ct-values became negative after dilution, leading to decreased sample sizes and less powerful statistics. For example, the number of AAV positive samples dropped from 23 out of 30 to 10.

Table 27: Summary of the normalisation (RT-)qPCRs. The numbers in brackets indicate the corresponding number before dilution

Virus	n healthy	Mean Ct healthy	n diarrhoea	Mean Ct diarrhoea	p-value
pKoV	15 (15)	26.25 (23.51)	14 (15)	26.40 (26.46)	0.90 (0.06)
pRVA	7 (11)	33.68 (30.72)	12 (13)	25.86 (21.89)	0.01 (0.005)
UBoV5	13 (15)	28.62 (22.75)	12 (14)	30.57 (26.14)	0.33 (0.12)
AAV	5 (10)	36.26 (29.73)	5 (13)	41.55 (37.08)	0.08 (0.005)
pPSV	6 (7)	31.94 (25.17)	2 (7)	31.92 (34.34)	0.99 (0.005)

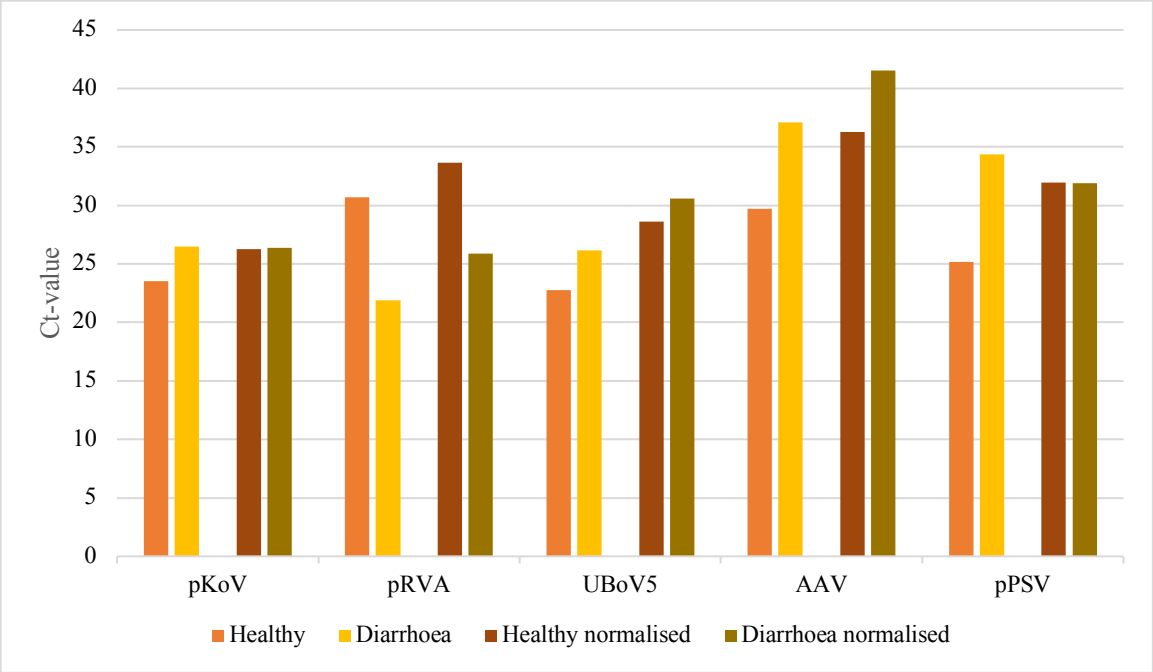

Figure 24: Mean Ct-values before and after normalisation

4 DISCUSSION

In this study, we analysed the faecal virome and bacteriome of 50 healthy and 50 diarrhoeic piglets to identify viruses that are part of the normal flora of young pigs and may interact in a commensal or even mutual way with their host. We analysed the faecal virome and bacteriome by metagenomic and 16S rRNA amplicon sequencing, respectively. Candidate viruses, which were associated with health, were then quantified by (RT-)qPCR.

4.1 High viral diversity in faecal samples of piglets

In total, we found 27 (porcine) viral genera (Table 22). This number is quite high, especially compared to previous metagenomic studies in pigs who found between nine to 15 different viruses (Shan et al., 2011, Lager et al., 2012, Sachsenröder et al., 2012, Zhang et al., 2014, Amimo et al., 2016, Karlsson et al., 2016). One reason for the high diversity in our study might be the increasing number of "novel" viruses present on GenBank and hence in databases used for screening NGS data. A good example for this phenomenon is the group of CRESS DNA viruses which are found increasingly in various species and environments. This group currently consists of six families that are thought to infect eukaryotic cells (Zhao et al., 2019). Their commonalities, as the name suggests, are the circular single-stranded DNA and the encoded protein Rep (replicase or replicase-like). Since 2014, the number of known CRESS DNA genera increased more than fivefold (Zhao et al., 2019), which demonstrates the increasing number of genomes in GenBank and might explain the high virus diversity detected in our study. For some of the CRESS DNA viruses, associations with diseases are known, for example for porcine Circovirus-2. Many, like the Smacoviridae, cannot be propagated in cell cultures so far, and their function and biology is largely unknown (Zhao et al., 2019).

Important epizootic viruses like PEDV and TGEV, which are major causes for porcine diarrhoea in other countries and are often detected in metagenomic studies, e.g. from China (Zhang et al., 2014), where porcine Coronaviruses are frequent, could not be found in our study. TGE is a reportable disease in Switzerland that the country is free from. For PEDV the status is not known but if it is present, it is extremely rare. The Swiss pig industry is quite isolated since imports of pigs or porcine material are strictly regulated. However, pKoV and Astrovirus for example were detected in our study, as well as in studies from other continents, which indicates that these viruses seem to be present in the whole world.

It is worth mentioning that we did not include bacteriophages in our analysis. Other studies have found reads of bacteriophages in variable quantities in healthy and diarrhoeic pigs (Shan et al., 2011, Sachsenröder et al., 2012, Zhang et al., 2014, Amimo et al., 2016, Theuns et al., 2018) and undoubtedly phages are an important part of the intestinal virome. However, we were primarily interested in eukaryotic viruses that may infect porcine cells and hence interact directly with the host. Analysis of the phageom was beyond the scope of this study but may still be investigated later, as the sequencing raw data contain also the phage sequences which could be compared, e.g., to the 16S data.

4.2 Higher viral diversity in healthy than diarrheic animals

We could show, that on average there were 1.79-times (0 – 9) more virus positive samples in healthy than diarrhoeic animals (Table 23). Furthermore, the viral diversity was higher in healthy compared to diarrheic animals (Figure 4). This phenomenon may be a consequence of a pure "wash-out" and dilution effect due to diarrhoea. The increased fluid influx into the intestine or the disturbed absorption during diarrhoea dilutes the enteric content, and the accelerated passage of faeces washes out the intestine, resulting in less viruses being shed per volume of faeces even if intestinal particle production remains the same. On the other side, the fact that viruses are more diverse and abundant in healthy animals may indeed be due to commensal viruses being an integral part of the normal flora of healthy animals and their absence more or less directly linked to diarrhoea. In diarrhoeic piglets, the whole enteric ecosystem is imbalanced and enterocytes, who are the likely place of replication for most eukaryotic enteric viruses, are often damaged or erased. This may also negatively influence the propagation of these viruses. It is difficult to determine which of the two scenarios reflects the true situation. A mixture of both, wash-out effect and commensal viruses as part of the normal flora, is probably most likely. In this study, we tried to overcome the wash-out effect by normalising our (RT-)qPCR results through dilution to mimic the higher fraction of fluid in diarrhoeic faeces (see chapter 3.3.7). Determination of the water fraction by drying of the samples may be another option but requires relatively high sample volumes, and the correlation of water fraction and Ct-value needs to be established for each (RT-)qPCR separately. To our knowledge, there is no reliable normalisation procedure for faecal samples available to date and hence the influence of the dilution and wash-out effect on viral diversity cannot be clearly determined.

In any case, we found more different virus genera in healthy than in diarrhoeic samples. Interestingly, other studies have found a lower viral diversity in healthy compared to diarrhoeic pigs (Shan et al., 2011, Lager et al., 2012, Zhang et al., 2014). However, many viruses we found associated with health, e.g. CRESS DNA viruses, have only been detected recently, and therefore were not included in the analysis of older studies. In contrary, a few years ago mainly rather pathogenic viruses of established virus families were known, which could explain the higher viral diversity in diarrhoeic animals in these studies. Another factor could be that the symbiotic status of viruses may vary, meaning that a virus could be commensal or even mutual in one animal, but become pathogenic in another. This depends on a lot of factors, like the health and immune status of the host, the husbandry condition, or even the age of the host and may explain different findings in different studies.

4.3 Viruses associated with health

Kobu- and members of the CRESS DNA viruses showed the strongest associations with health (Table 22). For pKoV, no clear association with diarrhoea has been shown. In some studies, the prevalence in healthy and diarrhoeic animals was similar (Verma et al., 2013, Chuchaona et al., 2017, Jackova et al., 2017). Other groups found either a slightly higher prevalence (Zhang et al., 2014, Valkó et al., 2019) or higher numbers of reads (Shan et al., 2011) in healthy animals. Some studies saw a higher prevalence of pKoV in suckling piglets compared to weaned animals (Reuter et al., 2009, Barry et al., 2011, Jackova et al., 2017, Valkó et al., 2019). In our study, the prevalence of pKoV was generally high and similar in healthy and diarrheic animals as well as in suckling and weaned piglets (regardless of the health status). The results of previous research and of our study support the hypothesis, that pKoV cannot clearly be assigned to diarrhoea, and together with the high prevalence all over the world, the possibility of it being an integral part of the normal flora, particularly of piglets under eight weeks of age, is high.

In contrast to pKoV, studies about CRESS DNA viruses in pigs are very sparse, with the exception of porcine Circovirus-2, which is a known pathogen. CRESS DNA viruses can be found ubiquitously, for example in diatoms (Tomaru et al., 2008), fungi (Yu et al., 2010), insects (Dayaram et al., 2015), soil (Reavy et al., 2015) and humans (Ng et al., 2015), and are increasingly found by metagenomic studies, also in pigs (Shan et al., 2011, Sachsenröder et al., 2012, Zhang et al., 2014, Karlsson et al., 2016). Again, in most cases their clinical meaning and biology is largely unknown. A new study proposes that human Smacoviridae, a member of the

group of CRESS DNA viruses, may actually infect Archaea rather than humans (Díez-Villaseñor and Rodriguez-Valera, 2019). If this was the case for pPSV too, it would not meet our inclusion criteria anymore, since the host is not the pig and the virus rather a phage than a eukaryotic virus. However, like the previously discussed bacteriophages, these viruses could still be (indirectly) associated with health by having a possible modulating effect on the intestinal ecosystem. Interestingly, pPSV was the only virus that was significantly negatively correlated with pRVA. This could be an indication that pPSV might have a protective effect for the host. This correlation has not been shown before. There are possible explanations for this phenomenon: Superinfection exclusion for example prohibits a second virus to either enter a cell or replicate in the presence of a first virus. One example for this is the interaction between infectious pancreatic necrosis virus (IPNV) and viral haemorrhagic septicaemia virus (VHSV), which both cause diseases in salmonids. Cell lines that were persistently infected with IPNV inhibited RNA synthesis of VHSV, and therefore the propagation of the "rival" was inhibited (Parreño et al., 2017). Another possibility could be the destruction of the target cells (possibly enterocytes) through pRVA, which are then no longer available for pPSV for replication. If the hypothesis of archaea as hosts of pPSV is true, another explanation might be that due to the diarrhoea caused by pRVA archaea become depleted, which would automatically also eliminate pPSV. Lastly, pPSV may stimulate the immune system, similar to what has been found in mice infected with Herpesviruses, that were resistant to subsequent bacterial infection through stimulation of the innate immune system (Barton et al., 2007).

The next candidate virus in our study was UBoV. This virus, belonging to the family of Parvoviridae (ICTV Master Species List 2018b.v2, MSL #34), was first described by a Swedish group in pigs with postweaning multisystemic wasting syndrome (Blomström et al., 2009). It has also been found previously in metagenomic studies of pigs in both healthy and diarrhoeic animals, but with a tendency towards diarrhoea (Shan et al., 2011, Sachsenröder et al., 2012, Lager et al., 2012, Zhang et al., 2014, Amimo et al., 2016). In our study, we found mainly two genetically different subtypes of the virus, namely UBoV2 and UBoV5. Quantification by qPCR showed a correlation with health only for UBoV5 with significantly lower Ct-values in healthy piglets. A different pathogenicity for the two subtypes could be one reason for the varying findings amongst previous studies.

AAV was the last of our candidates. This virus was only found in ten samples by NGS, but with a high association with health. Using qPCR, 67 animals were positive, however many with high

Ct-values, explaining the missing detection by metagenomic analysis (see chapter 3.3.5). AAV also belongs to the Parvoviridae family, but to the genus Dependoparvovirus. As the name suggests, these viruses need a helper virus for replication. AAV is used as a gene therapy vector, since it is reportedly apathogenic (Kotterman and Schaffer, 2014). This apathogenicity, coupled with a high prevalence in humans (Boutin et al., 2010), already indicates potential commensalism for AAV. It has not been described in pig faeces so far, but gene therapy studies using pigs as a model for humans have been successfully performed (Steines et al., 2016). The fact that AAV can reduce particle production and DNA replication of its helper virus could give an explanation why it may be commensal or even mutualistic (Timpe et al., 2006). Known helper viruses are Adeno-, Herpes- and Papillomaviruses. In our study, AAV was not associated with Adenovirus, and Herpes- and Papillomaviruses were not found in our samples. This indicates that there may be another helper virus in pigs. If this helper virus itself causes diarrhoea, infection with AAV may indeed be beneficial for the host, since AAV may reduce propagation of the helper virus.

4.4 Significant differences in the bacteriome of healthy and diarrhoeic animals

Concerning the bacterial diversity, we could show that all the analysed diversity indices were lower in diarrhoeic piglets (Figure 8). Species richness indicates how many different species there are. The Shannon index not only includes the number of different species, but also their abundance, which makes it a quantitative value. The effective richness is the exponent of the Shannon index, which gives the effective number of species present in the context of the Shannon index. The evenness finally indicates how equal the diversity is. So, if all species present have the same abundance, the evenness is maximal, but if one species is much more frequent than the others, the evenness is low. This means, that in our diarrhoeic samples, there were significantly fewer species present, that they were less abundant and also less even. Basically, this could be translated to the “overgrowth” of a few bacteria – or loss of the others – qualitatively and quantitatively. In humans, it has also been shown that the bacterial diversity is reduced in diarrhoea in children (Monira et al., 2012). In our study, Prevotellaceae and Ruminococcaceae were amongst the most abundant bacterial families in healthy piglets. A similar result has been shown in other studies (Mach et al., 2015, Ramayo-Caldas et al., 2016, Dou et al., 2017, Vo et al., 2017). In contrast, mainly Enterobacteriaceae and Fusobacteriaceae

were dominant in diarrhoeic piglets in our study. This is also supported by literature (Bin et al., 2018).

It was also visible that the viral as well as the bacterial diversity increased with age, independent of the health status (Figure 5, Figure 13). In humans, it was also found that the bacteriome expands with increasing age (Lim et al., 2015). In our study, especially Prevotellaceae and Veillonellaceae seemed to be more abundant in older piglets, whereas Fusobacteriaceae and Bacteroidaceae prevailed in younger piglets. A similar observation has been made in another study (Vo et al., 2017). The piglets gain bacteria and viruses and develop their intestinal ecosystem with increasing age. What also should be remembered is that the age range in our study includes the weaning process, which has a substantial influence on the composition of the enteric flora. This switch can be nicely seen in Figure 14, where the abundance of the bacterial families changes around four weeks of age, independent of the health status. While this switch affects to some degree the same bacterial families as have been shown to be associated with health or diarrhoea (e.g. Prevotellaceae, Fusobacteriaceae) we can still see a significant difference in abundance between healthy and diarrhoeic animals in all age groups (Figure 12). However, to draw final conclusions on the temporal dynamics of viral and bacterial infections, piglets would need to be followed through the first weeks of life and sampled at regular intervals which is not the case in our study.

4.5 Association between viruses and bacteria associated with health

We could show that the presence of our candidate viruses was positively correlated with the abundance of bacteria associated with health and negatively with bacteria associated with diarrhoea (Figure 15). To our knowledge, no study exists that directly compares associations between the virome and the bacteriome in pigs. That said, the consequences of feeding probiotics on the virome have indeed been researched (Sachsenröder et al., 2014), but no influence of the probiotic treatment could be shown. Our results therefore contribute to identifying potentially commensal viruses, but also to the general knowledge of the faecal ecosystem and its interactions. The components should not only be looked at individually, but also together. However, the present data do not allow us to draw any conclusions as to the nature of the association between commensal viruses and bacteria of the normal flora.

4.6 Normalisation attempt

The quantitative investigations revealed significantly lower Ct-values for pKoV, UBoV5, pPSV and AAV in healthy animals compared to diarrhoeic piglets (Figure 17, Figure 20, Figure 22, Figure 23). To our knowledge, those viruses have not been quantitatively compared by (RT-)qPCR between healthy and diarrhoeic pigs before. The (RT-)qPCR data confirmed not only the NGS results showing that healthy piglets are positive for more viruses but revealed also that they have higher viral loads than the diarrhoeic ones. Even though the quantitative differences were not statistically significant anymore when a subset of 30 samples was diluted to the same optical density in an attempt to normalise the different consistencies of faecal samples, the viral loads of our candidate viruses were still higher or equal in healthy animals compared to diarrhoeic animals (Figure 24). Hence, while we cannot proclaim that the candidate viruses are mutualistic, our data clearly suggest that they are not detrimental to the host and therefore classified as commensal. Notably, because of the massive dilution, in many samples the viruses became undetectable, especially if the Ct-values of the undiluted samples were already high. This led to a loss of number of positive samples and therefore a reduction of the sample size (e.g. from 26 positive samples to ten for AAV), which rendered the difference in mean Ct-values statistically non-significant even though the difference was still high. Testing a higher number of samples with the normalisation method would therefore be important. However, to dilute faeces of diverse consistency, colour and composition to the same optical density is rather tedious and time consuming as each sample needs to be handled individually and the method is therefore not well feasible for a larger number of samples. Another factor to keep in mind is that we simply diluted our samples with water to mimic the increased fluid content of diarrhoea. In reality however, diarrhoea is not just characterised by an increased influx of water. The whole enteric environment and composition is imbalanced and general conditions such as the pH value are changed.

4.7 Limitations and outlook

The limitations of our study were the relatively small number of positive samples for some viruses and lack of an established method for normalisation of the (RT-)qPCR-results. We could show that our candidates were associated with commensal bacteria, but we cannot conclude on the nature of this association. The same is true for the negative association of pPSV with pRVA. Further studies will have to screen for the most promising candidates such as pPSV

and AAV in larger populations, and questions regarding cell tropism or even the hosts of these viruses will have to be answered. A possible way could be in-situ hybridisation of the intestine of infected pigs. Furthermore, the viruses would need to be propagated in cell culture to examine their biology and the effect of co-infections more deeply. Finally, experimental infections of pigs might show, if some of our candidates are really commensal or even beneficial for the host. This distinction can vary, since a virus can be commensal under certain circumstances and mutualistic or pathogenic under others. A lot of factors come into play, be it the age or immune status of the pig, or the housing conditions, that may confer stress leading to immunosuppression, or the composition and interplay of the other members of the intestinal/faecal ecosystem.

4.8 Conclusion

In summary, our data showed an astonishing diversity in the faecal virome in healthy and diarrhoeic piglets up to eight weeks of age. While the composition and diversity of the virome as well as the bacteriome is significantly influenced by age, we still found significant differences in the composition and a significantly higher viral and bacterial diversity in healthy animals. Several viruses such as pKoV and members of the CRESS DNA group (pPSV, pSCV) as well as two members of the Parvoviridae family (UBoV5, AAV) were significantly associated with health: qualitatively by significantly higher numbers of positive animals as determined by metagenomic analysis, and quantitatively by significantly higher viral genome load in healthy animals analysed by specific (RT-)qPCR. While the normalisation trial has shown that part of this finding may be attributed to dilution or wash-out effects leading to lower numbers of viruses per volume of diarrhoeic faeces, these viruses are still very likely commensal for the host, as the association of the candidate viruses with bacteria of the normal flora indicates. However, the knowledge about these viruses will have to be greatly increased for determining their definite function. The question also arises, whether they are just markers for a healthy faecal ecosystem, or if they can actively contribute to the health of their host. This knowledge may help to prevent or even treat diarrhoeic diseases in piglets and possibly humans.

REFERENCES

AMIMO, J. O., EL ZOWALATY, M. E., GITHAE, D., WAMALWA, M., DJIKENG, A. & NASRALLAH, G. K. 2016. Metagenomic analysis demonstrates the diversity of the fecal virome in asymptomatic pigs in East Africa. *Arch Virol,* 161**,** 887-97.

ARISAWA, K., SOBUE, T., YOSHIMI, I., SODA, M., SHIRAHAMA, S., DOI, H., KATAMINE, S., SAITO, H. & URATA, M. 2003. Human T-lymphotropic virus type-I infection, survival and cancer risk in southwestern Japan: a prospective cohort study. *Cancer Causes Control,* 14**,** 889-96.

BARRY, A. F., RIBEIRO, J., ALFIERI, A. F., VAN DER POEL, W. H. & ALFIERI, A. A. 2011. First detection of kobuvirus in farm animals in Brazil and the Netherlands. *Infect Genet Evol,* 11**,** 1811-4.

BARTON, E. S., WHITE, D. W., CATHELYN, J. S., BRETT-MCCLELLAN, K. A., ENGLE, M., DIAMOND, M. S., MILLER, V. L. & VIRGIN, H. W. 2007. Herpesvirus latency confers symbiotic protection from bacterial infection. *Nature,* 447**,** 326-9.

BIN, P., TANG, Z., LIU, S., CHEN, S., XIA, Y., LIU, J., WU, H. & ZHU, G. 2018. Intestinal microbiota mediates Enterotoxigenic Escherichia coli-induced diarrhea in piglets. *BMC Vet Res,* 14**,** 385.

BLOMSTRÖM, A. L., BELÁK, S., FOSSUM, C., MCKILLEN, J., ALLAN, G., WALLGREN, P. & BERG, M. 2009. Detection of a novel porcine boca-like virus in the background of porcine circovirus type 2 induced postweaning multisystemic wasting syndrome. *Virus Res,* 146**,** 125-9.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-20.

BOUTIN, S., MONTEILHET, V., VERON, P., LEBORGNE, C., BENVENISTE, O., MONTUS, M. F. & MASURIER, C. 2010. Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors. *Hum Gene Ther,* 21**,** 704-12.

CHUCHAONA, W., KHAMRIN, P., YODMEEKLIN, A., KONGKAEW, A., VACHIRACHEWIN, R., KUMTHIP, K., USHIJIMA, H. & MANEEKARN, N. 2017. Detection and molecular characterization of porcine kobuvirus in piglets in 2009-2013 in northern Thailand. *Trop Anim Health Prod,* 49**,** 1077-1080.

DAYARAM, A., POTTER, K. A., PAILES, R., MARINOV, M., ROSENSTEIN, D. D. & VARSANI, A. 2015. Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (Insecta: Odonata) of Arizona and Oklahoma, USA. *Infect Genet Evol,* 30**,** 278-287.

DOU, S., GADONNA-WIDEHEM, P., ROME, V., HAMOUDI, D., RHAZI, L., LAKHAL, L., LARCHER, T., BAHI-JABER, N., PINON-QUINTANA, A., GUYONVARCH, A., HUËROU-LURON, I. L. & ABDENNEBI-NAJAR, L. 2017. Characterisation of Early-Life Fecal Microbiota in Susceptible and Healthy Pigs to Post-Weaning Diarrhoea. *PLoS One,* 12**,** e0169851.

DUFKOVA, L., SCIGALKOVA, I., MOUTELIKOVA, R., MALENOVSKA, H. & PRODELALOVA, J. 2013. Genetic diversity of porcine sapoviruses, kobuviruses, and astroviruses in asymptomatic pigs: an emerging new sapovirus GIII genotype. *Arch Virol,* 158**,** 549-58.

DÍEZ-VILLASEÑOR, C. & RODRIGUEZ-VALERA, F. 2019. CRISPR analysis suggests that small circular single-stranded DNA smacoviruses infect Archaea instead of humans. *Nat Commun,* 10**,** 294.

EDGAR, R. C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods,* 10**,** 996-8.

GIANG, H. H., VIET, T. Q., OGLE, B. & LINDBERG, J. E. 2012. Growth performance, digestibility, gut environment and health status in weaned piglets fed a diet supplemented with a complex of lactic acid bacteria alone or in combination with Bacillus subtilis and Saccharomyces boulardii. *Livestock Science,* 143**,** 132-141.

GRIFFITHS, P. 1999. Time to consider the concept of a commensal virus? *Rev Med Virol,* 9**,** 73-4.

JACKOVA, A., SLIZ, I., MANDELIK, R., SALAMUNOVA, S., NOVOTNY, J., KOLESAROVA, M., VLASAKOVA, M. & VILCEK, S. 2017. Porcine kobuvirus 1 in healthy and diarrheic pigs: Genetic detection and characterization of virus and co-infection with rotavirus A. *Infect Genet Evol,* 49**,** 73-77.

KARLSSON, O. E., LARSSON, J., HAYER, J., BERG, M. & JACOBSON, M. 2016. The Intestinal Eukaryotic Virome in Healthy and Diarrhoeic Neonatal Piglets. *PLoS One,* 11**,** e0151481.

KERNBAUER, E., DING, Y. & CADWELL, K. 2014. An enteric virus can replace the beneficial function of commensal bacteria. *Nature,* 516**,** 94-8.

KLINDWORTH, A., PRUESSE, E., SCHWEER, T., PEPLIES, J., QUAST, C., HORN, M. & GLÖCKNER, F. O. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res,* 41**,** e1.

KOTTERMAN, M. A. & SCHAFFER, D. V. 2014. Engineering adeno-associated viruses for clinical gene therapy. *Nat Rev Genet,* 15**,** 445-51.

LAGER, K. M., NG, T. F., BAYLES, D. O., ALT, D. P., DELWART, E. L. & CHEUNG, A. K. 2012. Diversity of viruses detected by deep sequencing in pigs from a common background. *J Vet Diagn Invest,* 24**,** 1177-9.

LEMON, J. 2006. Plotrix: a package in the red light district of R. R-News 6: 8–12.

LIM, E. S., ZHOU, Y., ZHAO, G., BAUER, I. K., DROIT, L., NDAO, I. M., WARNER, B. B., TARR, P. I., WANG, D. & HOLTZ, L. R. 2015. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat Med,* 21**,** 1228-34.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

MACH, N., BERRI, M., ESTELLÉ, J., LEVENEZ, F., LEMONNIER, G., DENIS, C., LEPLAT, J. J., CHEVALEYRE, C., BILLON, Y., DORÉ, J., ROGEL-GAILLARD, C. & LEPAGE, P. 2015. Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ Microbiol Rep,* 7**,** 554-69.

MAECHLER, M., ROUSSEEUW, P., STRUYF, A., HUBERT, M. & HORNIK, K. 2012. Cluster: cluster analysis basics and extensions. *R package version,* 1**,** 56.

MAGURRAN, A. E. 2004. *Measuring Biological Diversity*, Blackwell Science Ltd.

MALASAO, R., KHAMRIN, P., KUMTHIP, K., USHIJIMA, H. & MANEEKARN, N. 2018. Complete genome sequence analysis of rare G4P[6] rotavirus strains from human and pig reveals the evidence for interspecies transmission. *Infect Genet Evol,* 65**,** 357-368.

MENG, X. J. 2013. Porcine circovirus type 2 (PCV2): pathogenesis and interaction with the immune system. *Annu Rev Anim Biosci,* 1**,** 43-64.

METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet,* 11**,** 31-46.
MONIRA, S., SHABNAM, S. A., ALAM, N. H., ENDTZ, H. P., CRAVIOTO, A. & ALAM, M. 2012. 16S rRNA gene-targeted TTGE in determining diversity of gut microbiota during acute diarrhoea and convalescence. *J Health Popul Nutr,* 30**,** 250-6.
NETER, J. & WASSERMAN, W. 1974. *Applied Linear Statistical Models,* Homewood (Illinois, USA), Richard D. Irwin Inc.
NG, T. F., ZHANG, W., SACHSENRÖDER, J., KONDOV, N. O., DA COSTA, A. C., VEGA, E., HOLTZ, L. R., WU, G., WANG, D., STINE, C. O., ANTONIO, M., MULVANEY, U. S., MUENCH, M. O., DENG, X., AMBERT-BALAY, K., POTHIER, P., VINJÉ, J. & DELWART, E. 2015. A diverse group of small circular ssDNA viral genomes in human and non-human primate stools. *Virus Evol,* 1**,** vev017.
OPRIESSNIG, T. & HALBUR, P. G. 2012. Concurrent infections are important for expression of porcine circovirus associated disease. *Virus Res,* 164**,** 20-32.
PARREÑO, R., ALMAGRO, L., BELLÓ-PÉREZ, M., MEDINA-GALI, R. M., ESTEPA, A. & PEREZ, L. 2017. Restricted replication of viral hemorrhagic septicemia virus (VHSV) in a birnavirus-carrier cell culture. *Arch Virol,* 162**,** 1037-1041.
PETTY, T. J., CORDEY, S., PADIOLEAU, I., DOCQUIER, M., TURIN, L., PREYNAT-SEAUVE, O., ZDOBNOV, E. M. & KAISER, L. 2014. Comprehensive human virus screening using high-throughput sequencing with a user-friendly representation of bioinformatics analysis: a pilot study. *J Clin Microbiol,* 52**,** 3351-61.
PIELOU, E. C. 1975. *Ecological Diversity*, Wiley.
RAMAYO-CALDAS, Y., MACH, N., LEPAGE, P., LEVENEZ, F., DENIS, C., LEMONNIER, G., LEPLAT, J. J., BILLON, Y., BERRI, M., DORÉ, J., ROGEL-GAILLARD, C. & ESTELLÉ, J. 2016. Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *ISME J,* 10**,** 2973-2977.
REAVY, B., SWANSON, M. M., COCK, P. J., DAWSON, L., FREITAG, T. E., SINGH, B. K., TORRANCE, L., MUSHEGIAN, A. R. & TALIANSKY, M. 2015. Distinct circular single-stranded DNA viruses exist in different soil types. *Appl Environ Microbiol,* 81**,** 3934-45.
REUTER, G., BOLDIZSÁR, A., KISS, I. & PANKOVICS, P. 2008. Candidate new species of Kobuvirus in porcine hosts. *Emerg Infect Dis,* 14**,** 1968-70.
REUTER, G., BOLDIZSÁR, A. & PANKOVICS, P. 2009. Complete nucleotide and amino acid sequences and genetic organization of porcine kobuvirus, a member of a new species in the genus Kobuvirus, family Picornaviridae. *Arch Virol,* 154**,** 101-8.
ROOSSINCK, M. J. 2011. The good viruses: viral mutualistic symbioses. *Nat Rev Microbiol,* 9**,** 99-108.
SACHSENRÖDER, J., TWARDZIOK, S., HAMMERL, J. A., JANCZYK, P., WREDE, P., HERTWIG, S. & JOHNE, R. 2012. Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. *PLoS One,* 7**,** e34631.
SACHSENRÖDER, J., TWARDZIOK, S. O., SCHEUCH, M. & JOHNE, R. 2014. The general composition of the faecal virome of pigs depends on age, but not on feeding with a probiotic bacterium. *PLoS One,* 9**,** e88888.
SEGALÉS, J., ALLAN, G. M. & DOMINGO, M. 2005. Porcine circovirus diseases. *Anim Health Res Rev,* 6**,** 119-42.

SHAN, T., LI, L., SIMMONDS, P., WANG, C., MOESER, A. & DELWART, E. 2011. The fecal virome of pigs on a high-density farm. *J Virol,* 85**,** 11697-708.
STEINES, B., DICKEY, D. D., BERGEN, J., EXCOFFON, K. J., WEINSTEIN, J. R., LI, X., YAN, Z., ABOU ALAIWA, M. H., SHAH, V. S., BOUZEK, D. C., POWERS, L. S., GANSEMER, N. D., OSTEDGAARD, L. S., ENGELHARDT, J. F., STOLTZ, D. A., WELSH, M. J., SINN, P. L., SCHAFFER, D. V. & ZABNER, J. 2016. gene transfer with AAV improves early cystic fibrosis pig phenotypes. *JCI Insight,* 1**,** e88728.
THEPIGSITE.COM. 2019. *Diarrhoea or scours* [Online]. Available: https://thepigsite.com/disease-guide/diarrhoea-scours [Accessed 12 2019].
THEUNS, S., VANMECHELEN, B., BERNAERT, Q., DEBOUTTE, W., VANDENHOLE, M., BELLER, L., MATTHIJNSSENS, J., MAES, P. & NAUWYNCK, H. J. 2018. Nanopore sequencing as a revolutionary diagnostic tool for porcine viral enteric disease complexes identifies porcine kobuvirus as an important enteric virus. *Sci Rep,* 8**,** 9830.
TIMPE, J. M., VERRILL, K. C. & TREMPE, J. P. 2006. Effects of adeno-associated virus on adenovirus replication and gene expression during coinfection. *J Virol,* 80**,** 7807-15.
TOMARU, Y., SHIRAI, Y., SUZUKI, H., NAGUMO, T. & NAGASAKI, K. 2008. Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom Chaetoceros debilis. *Aquatic Microbial Ecology,* 50**,** 103-112.
VALKÓ, A., MAROSI, A., CSÁGOLA, A., FARKAS, R., RÓNAI, Z. & DÁN, Á. 2019. Frequency of diarrhoea-associated viruses in swine of various ages in Hungary. *Acta Vet Hung,* 67**,** 140-150.
VARSANI, A. & KRUPOVIC, M. 2018. Smacoviridae: a new family of animal-associated single-stranded DNA viruses. *Arch Virol,* 163**,** 2005-2015.
VERMA, H., MOR, S. K., ABDEL-GLIL, M. Y. & GOYAL, S. M. 2013. Identification and molecular characterization of porcine kobuvirus in U. S. swine. *Virus Genes,* 46**,** 551-3.
VO, N., TSAI, T. C., MAXWELL, C. & CARBONERO, F. 2017. Early exposure to agricultural soil accelerates the maturation of the early-life pig gut microbiota. *Anaerobe,* 45**,** 31-39.
YU, X., LI, B., FU, Y., JIANG, D., GHABRIAL, S. A., LI, G., PENG, Y., XIE, J., CHENG, J., HUANG, J. & YI, X. 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proc Natl Acad Sci U S A,* 107**,** 8387-92.
ZHANG, B., TANG, C., YUE, H., REN, Y. & SONG, Z. 2014. Viral metagenomics analysis demonstrates the diversity of viral flora in piglet diarrhoeic faeces in China. *J Gen Virol,* 95**,** 1603-11.
ZHAO, L., ROSARIO, K., BREITBART, M. & DUFFY, S. 2019. Eukaryotic Circular Rep-Encoding Single-Stranded DNA (CRESS DNA) Viruses: Ubiquitous Viruses With Small Genomes and a Diverse Host Range. *Adv Virus Res,* 103**,** 71-133.
ZHOU, Y., XU, J., WANG, W. L., SONG, S. W., ZHU, S. K., MENG, Q. F., YU, F., LI, C. P., LIU, N. & LUAN, W. M. 2018. A TaqMan-based real-time PCR assay for the detection of Ungulate bocaparvovirus 2. *J Virol Methods,* 261**,** 17-21.

ACKNOWLEDGEMENTS

This work was funded by the Forschungskredit of the University of Zurich (FK-17-063).

First, I would like to especially thank PD Dr. Claudia Bachofen for the opportunity to complete my thesis with her and the great supervision, support and patience.

I thank Prof. Dr. Matthias Schweizer for co-examining my thesis.

Thank you to Giuliana Rosato and the other assistants of the Institute for Veterinary Pathology of the University of Zurich for collecting my samples.
Also, thanks to Dr. Yvonne Masserey and her colleagues at the Schweinegesundheitsdienst Zürich-Ost for collecting my samples.

I thank Dr. Marc W. Schmid for the bioinformatic work and the figures.

Further, I'd like to thank Prof. Dr. Cornel Fraefel for the opportunity to work at his institute and his scientific inputs.

A big thank you to all of my group members, especially Karin, Jakub, Isabelle, Deborah, Isabelle, Tamara, Julia, Marco and Martina for all the great times in- and outside the lab, the friendships and the invaluable support.

Thank you to all the members of our institute for the nice atmosphere and all the good memories.

Lastly, I'd like to thank my family and Philippe for their support, patience and encouragement.

CURRICULUM VITAE

Vorname Name	Charlotte Isabelle Rickli
Geburtsdatum	28.10.1991
Geburtsort	Basel, Schweiz
Nationalität	Schweiz
Heimatort	Madiswil BE

August 1998 – Juni 2002	Primarschule, Basel, Schweiz
August 2002 – Juni 2005	Orientierungsschule, Basel, Schweiz
August 2005 – Juni 2010	Gymnasium, Basel, Schweiz
25. Juni 2010	Matura, Gymnasium am Münsterplatz, Basel, Schweiz
September 2010 – Januar 2017	Studium der Veterinärmedizin, Universität Zürich, Schweiz
26. Januar 2017	Staatsexamen med. vet., Universität Zürich, Schweiz
September 2017 – November 2019	Anfertigung der Dissertation unter Leitung von PD Dr. med. vet. Claudia Bachofen am Virologischen Institut der Vetsuisse-Fakultät Universität Zürich Direktor Prof. Dr. sc. nat. ETH Cornel Fraefel